



```

: APPLICANT: The Regents of the University of California
: TITLE OF INVENTION: Assays for Taste Receptor Cell Specific
: FILE REFERENCE: Ion Channel
: CURRENT APPLICATION NUMBER: US/10/026,188
: PRIOR FILING DATE: 2001-12-21
: PRIOR APPLICATION NUMBER: US 60/259,379
: NUMBER OF SEQ ID NOS: 8
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 3
: LENGTH: 249487
: TYPE: DNA
: ORGANISM: Mus musculus
: FEATURE:
: OTHER INFORMATION: mouse genomic region containing Itprcs
US-10-026-188-3

Query Match
Best Local Similarity 58.8%; Score 19.4; DB 9; Length 249487;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 2 GGGATCCATGCGGGAACAGTAACGAT 30
Db 198467 GTGATCCATGCCACAGACAGGAATGAT 198495

RESULT 3
US-09-925-301-233/c
: Sequence 233, Application US/09925301
: Patent No. US20020052308A1
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
: FILE REFERENCE: PA106
: CURRENT APPLICATION NUMBER: US/09/925,301
: PRIOR FILING DATE: 2001-08-10
: PRIOR APPLICATION NUMBER: PCT/US00/05882
: PRIOR FILING DATE: 2000-03-08
: PRIOR APPLICATION NUMBER: 60/124,270
: PRIOR FILING DATE: 1999-03-12
: NUMBER OF SEQ ID NOS: 1694
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 233
: LENGTH: 2081
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-925-301-233

Query Match
Best Local Similarity 58.2%; Score 19.2; DB 10; Length 2081;
Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 1 GGGGATCCATGCGGGAACAGTAACGATTA 32
Db 396 GGGGATCCATGCGGGAACAGTAACGATTA 365

RESULT 4
US-10-098-841-213/c
: Sequence 213, Application US/10098841
: Publication No. US20020197679A1
: GENERAL INFORMATION:
: APPLICANT: Tang, Y. Tom
: APPLICANT: Liu, Chenghua
: APPLICANT: Asundi, Vinod
: APPLICANT: Xu, Chongjun
: APPLICANT: Zhou, Ping
: APPLICANT: Ma, Yungling
: APPLICANT: Wang, Jian-Rui
: APPLICANT: Zhao, Qing A.
: APPLICANT: Ren, Feiyan
: APPLICANT: Chen, Rui-hong
```

```

: APPLICANT: Wang, Dunrui
: APPLICANT: Wang, Zhiwei
: APPLICANT: Wehrman, Tom
: APPLICANT: Zhang, Jie
: APPLICANT: Qian, Xiaohong B.
: APPLICANT: Dimanac, Radoje T.
: TITLE OF INVENTION: No. US20020197679A1 Nucleic Acids and
: FILE REFERENCE: Polypeptides
: CURRENT APPLICATION NUMBER: US/10/098,841
: PRIOR FILING DATE: 2002-03-13
: PRIOR APPLICATION NUMBER: 09/598,042
: PRIOR FILING DATE: 2000-06-20
: PRIOR APPLICATION NUMBER: 09/552,317
: PRIOR FILING DATE: 2000-04-25
: PRIOR APPLICATION NUMBER: 09/488,725
: PRIOR FILING DATE: 2000-01-21
: NUMBER OF SEQ ID NOS: 331
: SOFTWARE: PL_FL-genes Version 1.0
: SEQ ID NO 213
: LENGTH: 5201
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (847)..(3063)
US-10-098-841-213

Query Match
Best Local Similarity 58.2%; Score 19.2; DB 9; Length 5201;
Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
```

```
OY 2 GGGATCCATGCGGGAACAGTAACGATTAC 33
Db 4476 GGAGATCCAGCGAGTAAGTAAGTAAGTCTGAC 4445
```

```

RESULT 5
US-09-954-531-130/c
: Sequence 130, Application US/09954531
: Patent No. US20020165180A1
: GENERAL INFORMATION:
: APPLICANT: Weaver, Zoe
: TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C
: FILE REFERENCE: Gene Sets
: CURRENT APPLICATION NUMBER: US/09/954,531
: PRIOR FILING DATE: 2002-05-02
: PRIOR APPLICATION NUMBER: US/60/233,133
: PRIOR FILING DATE: 2000-09-18
: PRIOR APPLICATION NUMBER: US/60/234,009
: PRIOR FILING DATE: 2000-09-20
: PRIOR APPLICATION NUMBER: US/60/234,034
: PRIOR FILING DATE: 2000-09-20
: PRIOR APPLICATION NUMBER: US/60/234,509
: PRIOR FILING DATE: 2000-09-22
: PRIOR APPLICATION NUMBER: US/60/234,567
: PRIOR FILING DATE: 2000-09-22
: NUMBER OF SEQ ID NOS: 1392
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 130
: LENGTH: 5404
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-954-531-130
```

```

Query Match
Best Local Similarity 58.2%; Score 19.2; DB 9; Length 5404;
Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 2 GGGATCCATGCGGGAACAGTAACGATTAC 33
Db 3434 GGAGATCCAGCGAGTAAGTAAGTAAGTCTGAC 3403
```

RESULT 6  
US-09-954-531-351/c  
; Sequence 351, Application US/09954531  
; Patent No. US20020165180A1  
; GENERAL INFORMATION:  
; APPLICANT: Weaver, Zoe  
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cande  
; TITLE OF INVENTION: Gene Sets  
; FILE REFERENCE: 689290-77  
; CURRENT APPLICATION NUMBER: US/09/954,531  
; CURRENT FILING DATE: 2002-05-02  
; PRIOR APPLICATION NUMBER: US/60/233,133  
; PRIOR FILING DATE: 2000-09-18  
; PRIOR APPLICATION NUMBER: US/60/234,009  
; PRIOR FILING DATE: 2000-09-20  
; PRIOR APPLICATION NUMBER: US/60/234,034  
; PRIOR FILING DATE: 2000-09-20  
; PRIOR APPLICATION NUMBER: US/60/234,509  
; PRIOR FILING DATE: 2000-09-22  
; PRIOR APPLICATION NUMBER: US/60/234,567  
; PRIOR FILING DATE: 2000-09-22  
; NUMBER OF SEQ ID NOS: 1392  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 351  
; LENGTH: 5404  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-954-531-351

Query Match 58.2%; Score 19.2; DB 9; Length 5404;  
Best Local Similarity 75.0%; Pred. No. 21;  
Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 2 GGGATCCATGCGGAGACAGATGATAC 33  
Db 3434 GGGATCCATGCGGAGACAGATGATGATGAC 3403  
RESULT 7  
US-10-015-219-938/c  
; Sequence 938, Application US/10015219  
; Publication No. US20030008299A1  
; GENERAL INFORMATION:  
; APPLICANT: Algate, Paul A.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; FILE REFERENCE: 210121.493C1  
; CURRENT APPLICATION NUMBER: US/10/015,219  
; CURRENT FILING DATE: 2002-03-02  
; NUMBER OF SEQ ID NOS: 1739  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 938  
; LENGTH: 439  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-015-219-938

Query Match 57.6%; Score 19; DB 9; Length 439;  
Best Local Similarity 81.5%; Pred. No. 18;  
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 3 GGGATCCATGCGGAGACAGATGACGA 29  
Db 97 GGGATCCATGCTGAGACAGATGAGA 71

RESULT 8  
US-09-777-564-938/c  
; Sequence 938, Application US/09777564  
; Patent No. US2002022591A1  
; GENERAL INFORMATION:

; APPLICANT: Algate, Paul A.  
; APPLICANT: Mannion, Jane  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER  
; FILE REFERENCE: 210121.493  
; CURRENT APPLICATION NUMBER: US/09/777,564  
; CURRENT FILING DATE: 2001-02-05  
; NUMBER OF SEQ ID NOS: 1730  
; SOFTWARE: FastSeq for Window Version 4.0  
; SEQ ID NO 938  
; LENGTH: 439  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-777-564-938

Query Match 57.6%; Score 19; DB 10; Length 439;  
Best Local Similarity 81.5%; Pred. No. 18;  
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 3 GGGATCCATGCGGAGACAGATGACGA 29  
Db 97 GGGATCCATGCTGAGACAGATGAGA 71

RESULT 9  
US-09-989-545-9/c  
; Sequence 9, Application US/09989545  
; Patent No. US20020164697A1  
; GENERAL INFORMATION:  
; APPLICANT: Lehner, Sophie  
; APPLICANT: Manning, Stephen  
; APPLICANT: Coyle, Anthony J.  
; APPLICANT: Gutierrez-Ramos, Jose-Carlos  
; TITLE OF INVENTION: No. US20020164697A1el Th2-Specific Molecules and Uses Thereof  
; FILE REFERENCE: 5800-10B  
; CURRENT APPLICATION NUMBER: US/09/989,545  
; CURRENT FILING DATE: 2001-11-20  
; PRIOR APPLICATION NUMBER: 09/168,229  
; PRIOR FILING DATE: 1998-10-07  
; PRIOR APPLICATION NUMBER: 09/258,670  
; PRIOR FILING DATE: 1999-02-26  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 9  
; LENGTH: 3266  
; TYPE: DNA  
; ORGANISM: Mus sp.  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (40)..(642)  
; NAME/KEY: unsure  
; LOCATION: (1)..(3266)  
; OTHER INFORMATION: n= a, g, c, or t  
US-09-989-545-9

Query Match 57.6%; Score 19; DB 9; Length 3266;  
Best Local Similarity 81.5%; Pred. No. 24;  
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 GGGATCCATGCGGAGACAGATGACG 28  
Db 2366 GGGATCCATGCTGAGACAGATGAGA 2340

RESULT 10  
US-09-917-800A-1645/c  
; Sequence 1645, Application US/09917800A  
; Patent No. US20020119462A1  
; GENERAL INFORMATION:  
; APPLICANT: Mendrick, Donna  
; APPLICANT: Porter, Mark  
; APPLICANT: Johnson, Kory  
; APPLICANT: Castle, Arthur





;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Howrey & Simon  
;; STREET: 1299 Pennsylvania Avenue N.W.  
;; CITY: Washington  
;; STATE: DC  
;; COUNTRY: USA  
;; ZIP: 20004  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Diskette  
;; COMPUTER: IBM Compatible  
;; OPERATING SYSTEM: DOS  
;; SOFTWARE: FastSeq for Windows Version 2.0  
;;  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/962,527  
;; FILING DATE: 24-Sep-2001  
;; CLASSIFICATION: <Unknown>  
;;  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 09/037,751  
;; FILING DATE: 10-march-1998  
;;  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Halluin, Albert P  
;; REGISTRATION NUMBER: 25,277  
;; REFERENCE/DOCKET NUMBER: 00801.0140.999  
;; TELEPHONE: 650-463-8109  
;; TELEFAX: 650-463-8400  
;; TELEX: <Unknown>  
;;  
;; INFORMATION FOR SEQ ID NO: 3:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 6425 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: unknown  
;; MOLECULE TYPE: Genomic RNA  
;; SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
US-09-962-527-3  
Query Match 53.9%; Score 17.8; DB 9; Length 6425;  
Best Local Similarity 58.6%; Pred. No. 97;  
Matches 17; Conservative 5; Mismatches 7; Indels 0; Gaps 0;  
QY 5 GATCCATGCCGAGACAGTACGATTAC 33  
Db 3398 GGUCGAGCAGGACAGCAUAUAGCAUUAUC 3426  
RESULT 14  
US-09-962-527-2  
Sequence 2, Application US/09962527  
Publication No. US20030049813A1  
GENERAL INFORMATION:  
APPLICANT: GARGER, STEPHEN  
HOLTZ, R. BARRY  
MCULLOCH, MICHAEL  
TURPEN, THOMAS  
TITLE OF INVENTION: A PROCESS FOR ISOLATING AND  
PURIFYING VIRUSES SOLUBLE PROTEINS AND PEPTIDES  
FROM PLANT SOURCES  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Howrey & Simon  
STREET: 1299 Pennsylvania Avenue N.W.  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/962,527

;; FILING DATE: 24-Sep-2001  
;; CLASSIFICATION: <Unknown>  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 09/037,751  
;; FILING DATE: 10-march-1998  
;;  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Halluin, Albert P  
;; REGISTRATION NUMBER: 25,277  
;; REFERENCE/DOCKET NUMBER: 00801.0140.999  
;; TELEPHONE: 650-463-8109  
;; TELEFAX: 650-463-8400  
;; TELEX: <Unknown>  
;;  
;; INFORMATION FOR SEQ ID NO: 2:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 6439 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: unknown  
;; MOLECULE TYPE: Genomic RNA  
;; SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-962-527-2  
Query Match 53.9%; Score 17.8; DB 9; Length 6439;  
Best Local Similarity 58.6%; Pred. No. 97;  
Matches 17; Conservative 5; Mismatches 7; Indels 0; Gaps 0;  
QY 5 GATCCATGCCGAGACAGTACGATTAC 33  
Db 3398 GGUCGAGCAGGACAGCAUAUAGCAUUAUC 3426  
RESULT 15  
US-09-962-527-5  
Sequence 5, Application US/09962527  
Publication No. US20030049813A1  
GENERAL INFORMATION:  
APPLICANT: GARGER, STEPHEN  
HOLTZ, R. BARRY  
MCULLOCH, MICHAEL  
TURPEN, THOMAS  
TITLE OF INVENTION: A PROCESS FOR ISOLATING AND  
PURIFYING VIRUSES SOLUBLE PROTEINS AND PEPTIDES  
FROM PLANT SOURCES  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Howrey & Simon  
STREET: 1299 Pennsylvania Avenue N.W.  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/962,527  
FILING DATE: 24-Sep-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/037,751  
FILING DATE: 10-march-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Halluin, Albert P  
REGISTRATION NUMBER: 25,277  
REFERENCE/DOCKET NUMBER: 00801.0140.999  
TELEPHONE: 650-463-8109  
TELEFAX: 650-463-8400  
TELEX: <Unknown>  
INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:  
LENGTH: 6446 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: genomic RNA  
SEQUENCE DESCRIPTION: SEQ ID NO: 5:  
us-09-962-527-5

Query Match 53.9%; Score 17.8; DB 9; Length 6446;  
Best Local Similarity 58.6%; Pred. No. 97;  
Matches 17; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 5 GATCCATGCCGACAGAGCTAAGATTAC 33  
| : | : | | | | | | : | | : | |  
DB 3398 GGUCGAUCCAGACACACAUAAGCAUUDAC 3426

Search completed: March 23, 2003, 17:16:40  
Job time : 81.7097 secs

GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 23, 2003, 14:20:51 ; Search time 214.548 Seconds  
(without alignments)  
2491.051 Million cell updates/sec

Title: US-09-873-546-2

Perfect score: 33

Sequence: 1 gggggatccatgcgcgaacagagtaacgattac 33

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

otal number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estcom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	75.8	454	9	AA723009 zH30e05.s
2	25	75.8	470	9	A1497811 tB89f05.x
3	25	75.8	664	12	BG699685 602681538
4	22.6	68.5	466	10	BE019848 bB60e12.y
5	21.8	66.1	410	13	BI350717 fr37c06.y
6	21.8	66.1	467	12	BG892366 fp75d04.y

7	21.8	66.1	558	13	BM034832
8	21.8	66.1	573	13 <td>BM186599</td>	BM186599
9	21.8	66.1	626	10 <td>AM134357</td>	AM134357
10	21.8	66.1	628	13 <td>BM172422</td>	BM172422
11	21.8	66.1	632	13 <td>BI475251</td>	BI475251
12	21.8	66.1	645	13 <td>BI547693</td>	BI547693
13	21.8	66.1	702	10 <td>AM305720</td>	AM305720
14	21.8	66.1	709	12 <td>BG699614</td>	BG699614
15	21.8	66.1	786	13 <td>BI552244</td>	BI552244
16	21.8	66.1	801	12 <td>BG702913</td>	BG702913
17	21.8	66.1	893	14 <td>BQ922385</td>	BQ922385
18	21.8	66.1	1013	14 <td>BQ881423</td>	BQ881423
19	21	63.6	571	13 <td>BM283669</td>	BM283669
20	21	63.6	597	17 <td>AZ143597</td>	AZ143597
21	21	63.6	597	13 <td>BI594830</td>	BI594830
22	21	63.6	613	14 <td>BM655049</td>	BM655049
23	20.2	61.2	174	9 <td>AA341050</td>	AA341050
24	20.2	61.2	309	14 <td>F06359</td>	F06359
25	20.2	61.2	312	10 <td>AM249313</td>	AM249313
26	20.2	61.2	320	14 <td>F07725</td>	F07725
27	20.2	61.2	336	9 <td>AA309505</td>	AA309505
28	20.2	61.2	355	10 <td>BE394856</td>	BE394856
29	20.2	61.2	359	14 <td>F06503</td>	F06503
30	20.2	61.2	364	14 <td>F06857</td>	F06857
31	20.2	61.2	372	9 <td>AA172156</td>	AA172156
32	20.2	61.2	396	10 <td>AV684237</td>	AV684237
33	20.2	61.2	408	10 <td>BE266662</td>	BE266662
34	20.2	61.2	410	9 <td>A1400668</td>	A1400668
35	20.2	61.2	415	10 <td>AV736264</td>	AV736264
36	20.2	61.2	417	9 <td>AA312757</td>	AA312757
37	20.2	61.2	426	14 <td>T62496</td>	T62496
38	20.2	61.2	450	9 <td>AA232077</td>	AA232077
39	20.2	61.2	462	12 <td>BF193212</td>	BF193212
40	20.2	61.2	467	10 <td>AV715589</td>	AV715589
41	20.2	61.2	470	14 <td>RI5962</td>	RI5962
42	20.2	61.2	493	9 <td>AL711183</td>	AL711183
43	20.2	61.2	515	14 <td>BM837087</td>	BM837087
44	20.2	61.2	524	14 <td>BM846631</td>	BM846631
45	20.2	61.2	536	9 <td>AL561640</td>	AL561640

## ALIGNMENTS

RESULT 1  
AA723009 454 bp mRNA linear EST 02-JAN-1998  
LOCUS zH30e05.s1 Soares\_pinea1\_gland\_N3HPG Homo sapiens cDNA clone  
DEFINITION IMAGE:413600 3', mRNA sequence.  
ACCESSION AA723009  
VERSION AA723009.1 GI:2740716  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 454)  
AUTHORS Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Maria,M., Martin  
'J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,  
White,Y., Wylie,T., Waterston,R. and Wilson,R.  
WashU-NCI human EST Project  
TITLE Unpublished (1997)  
JOURNAL  
COMMENT Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewatson.wustl.edu  
This clone is available royalty-free through LNC ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -40m13 fwd. ET from Amerisham  
High quality sequence stop: 453.



[illegible]

TITLE  
JOURNAL  
COMMENT

Washu Zebrafish EST Project 1998  
Unpublished (1998)  
Contact: Stephen L. Johnson  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: zbrafish@watson.wustl.edu  
cDNA Library Preparation: John Neai, cDNA Library Arrayed by:  
Matthew Clair, DNA Sequencing by: Washington University Genome  
Sequencing Center Clone distribution: Genome Systems, St. Louis,  
Missouri (web address: www.genomesystems.com) (email contact:  
info@genomesystems.com) and Research Genetics, Huntsville, Alabama  
(web address: www.resgen.com) (email contact: info@resgen.com) and  
ResourcenzentrumPrimateDatenbank, Berlin, Germany (web address:  
www.irtpd.de)  
Seq primer: T7  
High quality sequence stop: 288.  
Location/Qualifiers  
1..410  
/organism="Danio rerio"  
/db\_xref="taxon:7955"  
/clone="4955530"  
/clone\_1lb="zebrafish adult brain"  
/sex="mixed male and female"  
/tissue-type="brain"  
/dev\_stage="adult"  
/lab\_host="E. coli DH10B"  
/note="Vector: pZIRPLOC; Site\_1: NotI; Site\_2: SalI;  
Original library was constructed in lambdaZIRPLOC. Mass  
excision of the cDNA library was performed to yield  
pZIRPLOC plasmids. Insert check was done in original  
library."  
BASE COUNT      116 a      98 c      101 g      92 t      3 others  
ORIGIN

Query Match                  66.1%; Score 21.8; DB 13; Length 410;  
Best Local Similarity 92.0%; Pred. No. 55;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY    9 CATGCCGACAGACTAAGATTAC 33  
     ||||| G ||||| |||||||  
Db    230 CATGCCAGACGACAGTAACTTAC 254

RESULT 6  
Bg892366

LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

Bg892366                  467 bp      mRNA      linear      EST 13-FEB-2002  
fp75d04.y1 zebrafish adult retina cDNA Danio rerio cDNA clone  
4787766.5 similar to TR:095057 095057 BC41195.1.; mRNA sequence.  
Bg892366                  Bg892366.1    GI:14286976  
EST.  
zebrafish.  
Danio rerio  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
Cyprinidae; Danio.  
1 (bases 1 to 467)  
Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Maria,M., Eddy  
S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood  
K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,  
Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,  
Kohn,S., Shih,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.  
and Wilson,R.  
Washu zebrafish EST Project 1998  
Unpublished (1998)  
Contact: Stephen L. Johnson  
Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: zbrafish@watson.wustl.edu

Library constructed by: Chandra Tucker and Gregory Niemi DNA  
Sequencing by: Washington University Genome Sequencing Center Clome  
distribution: RessourcenzentrumPrimatdatenbank, Berlin, Germany  
(web address: www.rzpd.de)  
High quality sequence stop: 390.

# FEATURES

## Source

Location/Qualifiers

1..467

/organism="Danio rerio"

/strain="wild-type"

/db\_xref="taxon:7955"

/clone="4787766"

/clone\_lib="zebrafish adult retina cdna"

/sex="mixed"

/dev\_stage="1-2 years"

/lab\_host="E.coli XL1-Blue MRF' (XL1-Blue MRF')"

/note="Vector: lambda ZAP II (pluscript SK-); Site\_1:

EcoRI; Site\_2: SalI; This zebrafish library was

constructed by Dr. Susan E. Brockhoff (email:

sbrockhoff@u.washington.edu) RZPD library number: 760"

BASE COUNT 125 a 124 c 118 g 99 t 1 others

ORIGIN

Query Match 66.1%; Score 21.8; DB 12; Length 467;

Best Local Similarity 92.0%; Pred. No. 56;

Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 188 CATGCCAGACAGATACGATTAC 212

QY 9 CATGCCAGACAGATACGATTAC 33

||||| |||||||

LOCUS 188 CATGCCAGACAGATACGATTAC 212

DEFINITION 558 bp mRNA linear EST 26-JUL-2002

similar to TR:095057 095057 BC41195.1.; mRNA sequence.

ACCESSION BM034832

VERSION BM034832.1 GI:16748403

KEYWORDS EST.

SOURCE zebrafish.

ORGANISM Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes

; Cyprinidae; Danio.

REFERENCE 1 (bases 1 to 558)

Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy

, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood

, K., Stepoe, M., Theising, B., Allen, M., Bowers, Y., Person, B.,

Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E.,

Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R.

and Wilson, R.

Washu zebrafish EST Project 1998

Unpublished (1998)

Other-ESTs: f36f03.x1

Contact: Stephen L. Johnson

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: zbrafish@watson.wustl.edu

CDNA Library Preparation: John Ngai. CDNA Library Arrayed by:

Matthew Clark. DNA Sequencing by: Washington University Genome

Sequencing Center Clome distribution: Genome Systems, St. Louis,

Missouri (web address: www.genomesystems.com) (email contact:

info@genomesystems.com) and Research Genetics, Huntsville, Alabama

(web address: www.resgen.com) (email contact: info@resgen.com) and

RessourcenzentrumPrimatdatenbank, Berlin, Germany (web address:

www.rzpd.de)

Seq primer: T7

High quality sequence stop: 437.

Location/Qualifiers

1..558

/organism="Danio rerio"

/db\_xref="taxon:7955"

/clone="5331604"

/clone\_lib="zebrafish adult brain"

/sex="mixed male and female"

/tissue\_type="brain"

/dev\_stage="adult"

/lab\_host="E. coli DH10B"

/note="Vector: pZiPlox; Site\_1: NotI; Site\_2: SalI;

Original library was constructed in lambdaZiPlox. Mass

excision of the cdna library was performed to yield

pZiPlox plasmids. Insert check was done in original

library."

BASE COUNT 150 a 153 c 137 g 118 t

ORIGIN

Query Match 66.1%; Score 21.8; DB 13; Length 558;

Best Local Similarity 92.0%; Pred. No. 58;

Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 228 CATGCCAGACAGATACGATTAC 252

QY 9 CATGCCAGACAGATACGATTAC 33

||||| |||||||

LOCUS BM186599

DEFINITION 573 bp mRNA linear EST 26-JUL-2002

similar to TR:095057 095057 BC41195.1.; mRNA sequence.

ACCESSION BM186599

VERSION BM186599.1 GI:17517557

KEYWORDS EST.

SOURCE zebrafish.

ORGANISM Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes

; Cyprinidae; Danio.

REFERENCE 1 (bases 1 to 573)

Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy

, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood

, K., Stepoe, M., Theising, B., Allen, M., Bowers, Y., Person, B.,

Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E.,

Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R.

and Wilson, R.

Washu zebrafish EST Project 1998

Unpublished (1998)

Other-ESTs: f36f03.x1

Contact: Stephen L. Johnson

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: zbrafish@watson.wustl.edu

CDNA Library Preparation: John Ngai. CDNA Library Arrayed by:

Matthew Clark. DNA Sequencing by: Washington University Genome

Sequencing Center Clome distribution: Genome Systems, St. Louis,

Missouri (web address: www.genomesystems.com) (email contact:

info@genomesystems.com) and Research Genetics, Huntsville, Alabama

(web address: www.resgen.com) (email contact: info@resgen.com) and

RessourcenzentrumPrimatdatenbank, Berlin, Germany (web address:

www.rzpd.de)

Seq primer: T7

High quality sequence stop: 449.

Location/Qualifiers

1..573

/organism="Danio rerio"

/db\_xref="taxon:7955"

/clone="5414284"

/clone\_lib="zebrafish adult brain"

/sex="mixed male and female"

```

/tissue_type="brain"
/dev_stage="adult"
/lab_host="E. coli DH10B"
/notes="Vector: pZIRPlox, Site.1: NotI; Site.2: SalI;
Original library was constructed in lambdaZIRPlox. Mass
excision of the cDNA library was performed to yield
pZIRPlox plasmids. Insert check was done in original
library."
BASE COUNT      151 a      157 c      144 g      121 t
ORIGIN
Query Match      66.1%; Score 21.8; DB 13; Length 573;
Best Local Similarity 92.0%; Pred. No. 58;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 9 CATGCCGAGACAGATTAACGATTAC 33
||||| || ||||| ||||| |||||
Db 228 CATGCCGAGACAGATTAACGATTAC 252

RESULT 9
AM134357 626 bp mRNA linear EST 27-OCT-1999
LOCUS f118d11.y1 Sugano Kawakami zebrafish DRA Danio rerio cDNA clone
DEFINITION 2601525 5' similar to TR:095057 095057 BC41195.1.; mRNA sequence.
ACCESSION AM134357
VERSION AM134357.1 GI:6135964
KEYWORDS EST.
SOURCE zebrafish.
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
; Cyprinidae; Danio.
1 (bases 1 to 626)
Sugano, S., Kawakami, K., Johnson, S., Li, F., Marra, M., Eddy, S.,
Hillier, L., Clifton, S., Allen, M., Gibbons, M., Jost, S., Kucaba, T.,
Martin, J., Pape, D., Steptoe, M., Underwood, K., Theising, B., Riltter
, E., Bowers, Y., Wylie, T., Waterston, R. and Wilson, R.
Washu Zebrafish EST Project 1999
Unpublished (1999)
Other_ESTS: f118d11.x1
Contact: S.L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
Sequencing by: Washington University Genome Sequencing Center
Seq primer: T3 ET from Amersham
High quality sequence stop: 509.
Location/Qualifiers
1. 626
/organism="Danio rerio"
/strain="AB"
/db_xref="taxon:7955"
/clone_id="2601525"
/clone_lib="Sugano Kawakami zebrafish DRA"
/sex="mixed (one male and one female, including
unfertilized eggs)"
/dev_stage="adult"
/lab_host="DH10B (phage resistant)"
/notes="Vector: pME18S-FL3; Site.1: DraIII (CACCCTGTG);
Site.2: DraIII (CACCCTGTG); 1st strand cDNA was primed
with an oligo(dT) primer [ATGTGGCTTTTCTTTTCTTTT];
double-stranded cDNA was ligated to a DraIII adaptor
[TCGTGGCTTCTG], digested and cloned into distinct DraIII
sites of the pME18S-FL3 vector (5' site CACCTGTG, 3' site
CACCCTGTG). XhoI should be used to isolate the cDNA
insert. Size selection was performed to exclude fragments
<1.5kb. Library constructed by Dr. Sumio Sugano
(University of Tokyo Institute of Medical Science) and
kindly donated by Dr. Koichi Kawakami. Custom primers for

```

```

sequencing: 5' end primer CTCTGCTCTTAAGACTCGC and 3' end
primer CGACCTGCAGCTCAGACAC.
BASE COUNT      168 a      166 c      167 g      123 t      2 others
ORIGIN
Query Match      66.1%; Score 21.8; DB 10; Length 626;
Best Local Similarity 92.0%; Pred. No. 59;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 9 CATGCCGAGACAGATTAACGATTAC 33
||||| || ||||| ||||| |||||
Db 209 CATGCCGAGACAGATTAACGATTAC 233

RESULT 10
BM172422 628 bp mRNA linear EST 04-DEC-2001
LOCUS imageq4_4-2001/smu47bdf41.x1 NIH_MGC_95 Homo sapiens cDNA clone
DEFINITION IMAGE:4816673 5', mRNA sequence.
ACCESSION BM172422
VERSION BM172422.1 GI:17311985
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 628)
Kale, P.I., Harsch, T.J., Folta, P.A., Nelson, D.O., Sanders, C.G. and
Prange, C.K.
The I.M.A.G.E. Consortium quality control effort: clone
resequencing for verification
Unpublished (2001)
Contact: Prange CK
The I.M.A.G.E. Consortium
Lawrence Livermore National Laboratory
Livermore, CA, USA
Email: help@image.llnl.gov
This read has been produced as part of the I.M.A.G.E. Consortium
quality control effort. High quality sequence is defined as having
100 or more base pairs with a phred quality value of 20 or greater,
where a sliding window of 4 base pairs with a phred quality value
of 15 or greater marks the beginning and end of the sequence. For
information on obtaining this clone, please contact
info@image.llnl.gov.
Plate: ILAM10715 row: 1 column: 18
Seq primer: -21m13
High quality sequence stop: 628.
Location/Qualifiers
1. 628
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image:4816673"
/clone_id="NIH_MGC_95"
/tissue_type="hippocampus"
/lab_host="DH10B"
/notes="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site.1: BamHI; Site.2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size selected for average insert size 2.5 kb and
normalized to ROP 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NIHRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      139 a      175 c      197 g      116 t      1 others
ORIGIN
Query Match      66.1%; Score 21.8; DB 13; Length 628;
Best Local Similarity 92.0%; Pred. No. 59;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 9 CATGCCGAGACAGATTAACGATTAC 33
||||| || ||||| ||||| |||||

```

DB 129 CATGCTGATCAGACTACGATTAC 153

RESULT 11  
BI475251

LOCUS  
DEFINITION f1q3a09.y3 zebrafish adult brain Dantio rerio cDNA clone 4832920 5' similar to tr:095057 095057 BC41195\_1. ; mRNA sequence.

ACCESSION  
VERSION BI475251  
KEYWORDS BI475251.1 GI:15306381

SOURCE  
ORGANISM zebrafish.  
Dantio rerio

REFERENCE  
AUTHORS Clark, M., Johnson, S. L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Stepien, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Rittler, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R., and Wilson, R.

TITLE Washu zebrafish EST Project 1998

JOURNAL  
COMMENT Unpublished (1998)  
Other\_Est: f1q3a09.x3  
Contact: Stephen L. Johnson  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: zbrafish@wustl.wustl.edu

FEATURES  
source  
1.632  
/organism="Dantio rerio"  
/db\_xref="taxon:7955"  
/clone="4832920"  
/clone\_lib="zebrafish adult brain"  
/sex="mixed male and female"  
/tissue\_type="brain"  
/dev\_stage="adult"  
/lab\_host="E. coli DH10B"  
/note="Vector: pZIRPlox; Site\_1: NotI; Site\_2: SalI;  
Original library was constructed in JambdazirPlox. Mass  
excision of the cDNA library was performed to yield  
pZIRPlox plasmids. Insert check was done in original  
library."

BASE COUNT 172 a 169 c 164 g 127 t

ORIGIN

Query Match 66.1%; Score 21.8; DB 13; Length 632;  
Best Local Similarity 92.0%; Pred. No. 59;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

LOCUS  
DEFINITION f1q3a09.y3 zebrafish adult brain Dantio rerio cDNA 5' similar to  
tr:095057 095057 BC41195\_1. ; mRNA sequence.

ACCESSION  
VERSION AM305720  
KEYWORDS AM305720.1 GI:6718073

SOURCE  
ORGANISM zebrafish.  
Dantio rerio

REFERENCE  
AUTHORS Clark, M., Johnson, S. L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Stepien, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Rittler, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R., and Wilson, R.

TITLE Washu zebrafish EST Project 1998

JOURNAL  
COMMENT Unpublished (1998)  
Other\_Est: f1q3a09.x3  
Contact: Stephen L. Johnson  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: zbrafish@wustl.wustl.edu

FEATURES  
source  
1.645  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5262995"  
/clone\_lib="NIH\_MGC\_95"  
/tissue\_type="hippocampus"  
/lab\_host="DH10B"  
/note="Organ: brain; Vector: pBluescript (modified  
pBluescript KS+); Site\_1: BamHI; Site\_2: SalI-XhoI (gtcgag  
); Oligo-dt primed using primer 5'-TTTTTTTTTTTTTTVN-3',  
size-selected for average insert size 2.5 kb and  
normalized to ROT 5. This is a primary library enriched  
for full-length clones and constructed using the  
Cap-trapper method (Carninci, in preparation). Library  
constructed by M. Brownstein (NIMH/NHGRI, National  
Institutes of Health). Note: this is a NIH\_MGC Library."

BASE COUNT 161 a 166 c 205 g 113 t

ORIGIN

Query Match 66.1%; Score 21.8; DB 13; Length 645;  
Best Local Similarity 92.0%; Pred. No. 59;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

LOCUS  
DEFINITION f1q3a09.y1 zebrafish adult brain Dantio rerio cDNA 5' similar to  
tr:095057 095057 BC41195\_1. ; mRNA sequence.

ACCESSION  
VERSION AM305720  
KEYWORDS AM305720.1 GI:6718073

SOURCE  
ORGANISM zebrafish.  
Dantio rerio

REFERENCE  
AUTHORS Clark, M., Johnson, S. L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Stepien, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Rittler, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R., and Wilson, R.

ACCESSION  
VERSION BI547693  
KEYWORDS BI547693.1 GI:15435005

SOURCE  
ORGANISM human.  
Homo sapiens

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
NIH-MGC http://mgc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)

JOURNAL  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shireki  
Toshiyuki and Piero Carninci (RIKEN)

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLM11663 row: a column: 12

High quality sequence stop: 634.

FEATURES  
source  
1.645  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5262995"  
/clone\_lib="NIH\_MGC\_95"  
/tissue\_type="hippocampus"  
/lab\_host="DH10B"  
/note="Organ: brain; Vector: pBluescript (modified  
pBluescript KS+); Site\_1: BamHI; Site\_2: SalI-XhoI (gtcgag  
); Oligo-dt primed using primer 5'-TTTTTTTTTTTTTTVN-3',  
size-selected for average insert size 2.5 kb and  
normalized to ROT 5. This is a primary library enriched  
for full-length clones and constructed using the  
Cap-trapper method (Carninci, in preparation). Library  
constructed by M. Brownstein (NIMH/NHGRI, National  
Institutes of Health). Note: this is a NIH\_MGC Library."

BASE COUNT 161 a 166 c 205 g 113 t

ORIGIN

Query Match 66.1%; Score 21.8; DB 13; Length 645;  
Best Local Similarity 92.0%; Pred. No. 59;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

LOCUS  
DEFINITION f1q3a09.y1 zebrafish adult brain Dantio rerio cDNA 5' similar to  
tr:095057 095057 BC41195\_1. ; mRNA sequence.

ACCESSION  
VERSION AM305720  
KEYWORDS AM305720.1 GI:6718073

SOURCE  
ORGANISM zebrafish.  
Dantio rerio

REFERENCE  
AUTHORS Clark, M., Johnson, S. L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Stepien, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Rittler, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R., and Wilson, R.



TITLE  
JOURNAL  
COMMENT

Washu Zebrafish EST Project 1998  
Unpublished (1998)  
Contact: Stephen L. Johnson  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: zbrafish@wustl.edu

CDNA Library Preparation: John Ngai. CDNA Library Arrayed by:  
Matthew Clark. DNA Sequencing by: Washington University Genome  
Sequencing Center Clone distribution: Genome Systems, St. Louis,  
Missouri (web address: www.genomesystems.com) (email contact:  
info@genomesystems.com) and Research Genetics, Huntsville, Alabama  
(web address: www.resgen.com) (email contact: info@resgen.com) and  
ResourceCenter@primatdatenbank, Berlin, Germany (web address:  
www.rzpd.de) '77  
Seq primer: '77  
High quality sequence stop: 486.

# FEATURES

source  
Location/Qualifiers

1. .702  
/organism="Danio rerio"  
/db\_xref="taxon:7955"  
/clone\_lib="zebrafish adult brain"  
/sex="mixed male and female"  
/tissue\_type="brain"  
/dev\_stage="adult"  
/lab\_host="E. coli DH10B"  
/note="Vector: pZIRPlox; Site.1: NotI; Site.2: SalI;  
original library was constructed in lambdaZIRPlox. Mass  
excision of the cDNA library was performed to yield  
pZIRPlox plasmids. Insert check was done in original  
library."  
BASE COUNT 194 a 189 c 181 g 137 t 1 others  
ORIGIN

Query Match 66.1%; Score 21.8; DB 10; Length 702;  
Best Local Similarity 92.0%; Pred. No. 60;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 9 CATGCCGACAGATACGATTAC 33  
||||| |||||||  
Db 219 CATGCCAGACAGATACGATTAC 243

# RESULT 14

LOCUS BG699614 709 bp mRNA linear EST 07-MAY-2001  
DEFINITION 60267934F1 NIH\_MGC\_95 Homo sapiens cDNA clone IMAGE:4811962 5',  
mRNA sequence.  
ACCESSION BG699614  
VERSION BG699614.1 GI:13968106  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

# REFERENCE

1 (bases 1 to 709)  
NIH-MGC http://mgi.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Miklos Palokovits, M.D., Ph.D.  
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki  
Toshiyuki and Piero Carninci (RIKEN)

# JOURNAL

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.jnl.gov  
Plate: LLM10703 row: h column: 11  
High quality sequence stop: 701.  
Location/Qualifiers

# FEATURES

source  
Location/Qualifiers

# source

1. .709  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4811962"  
/clone\_lib="NIH\_MGC\_95"  
/tissue\_type="hippocampus"  
/lab\_host="DH10B"  
/note="Organ: brain; Vector: pBluescriptR (modified  
pBluescript KS+); Site.1: BamHI; Site.2: SalI-XhoI (gtcgag  
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',  
size-selected for average insert size 2.5 kb and  
normalized to R0T 5. This is a primary library enriched  
for full-length clones and constructed using the  
Cap-trapper method (Carninci, in preparation). Library  
constructed by M. Brownstein (NIH/NHGRI, National  
Institutes of Health). Note: this is a NIH\_MGC Library."  
BASE COUNT 168 a 191 c 226 g 124 t  
ORIGIN

Query Match 66.1%; Score 21.8; DB 12; Length 709;  
Best Local Similarity 92.0%; Pred. No. 60;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 9 CATGCCGACAGATACGATTAC 33  
||||| |||||||  
Db 150 CATGCCAGACAGATACGATTAC 174

# RESULT 15

LOCUS BI552244 786 bp mRNA linear EST 05-SEP-2001  
DEFINITION 603195901F1 NIH\_MGC\_95 Homo sapiens cDNA clone IMAGE:5275297 5',  
mRNA sequence.  
ACCESSION BI552244  
VERSION BI552244.1 GI:15439556  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

Query Match 66.1%; Score 21.8; DB 12; Length 709;  
Best Local Similarity 92.0%; Pred. No. 60;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 9 CATGCCGACAGATACGATTAC 33  
||||| |||||||  
Db 150 CATGCCAGACAGATACGATTAC 174

# REFERENCE

1 (bases 1 to 786)  
NIH-MGC http://mgi.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Miklos Palokovits, M.D., Ph.D.  
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki  
Toshiyuki and Piero Carninci (RIKEN)

# JOURNAL

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.jnl.gov  
Plate: LLM11695 row: b column: 02  
High quality sequence stop: 589.  
Location/Qualifiers

# FEATURES

source  
Location/Qualifiers

1. .786  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5275297"  
/clone\_lib="NIH\_MGC\_95"  
/tissue\_type="hippocampus"  
/lab\_host="DH10B"  
/note="Organ: brain; Vector: pBluescriptR (modified  
pBluescript KS+); Site.1: BamHI; Site.2: SalI-XhoI (gtcgag  
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',  
size-selected for average insert size 2.5 kb and  
normalized to R0T 5. This is a primary library enriched  
for full-length clones and constructed using the  
Cap-trapper method (Carninci, in preparation). Library  
constructed by M. Brownstein (NIH/NHGRI, National  
Institutes of Health). Note: this is a NIH\_MGC Library."  
BASE COUNT 168 a 191 c 226 g 124 t  
ORIGIN

BASE COUNT 182 a 228 c 238 g 138 t  
ORIGIN

Query Match 66.1%; Score 21.8; DB 13; Length 786;  
Best Local Similarity 92.0%; Pred. No. 61;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 9 CATGCCGACAGAGTAACGATTAC 33  
||||| |||||||  
Db 73 CATGCCGACAGAGTAACGATTAC 97

Search completed: March 23, 2003, 17:12:20  
Job time : 220.548 secs

GenCore version 5.1.4\_p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 23, 2003, 13:56:56 ; Search time 113.806 Seconds  
(without alignments)  
8438.824 Million cell updates/sec

Title: US-09-873-546-2  
Perfect score: 33  
Sequence: 1 ggggagtcacatccggagacagagtaacgattac 33

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenBankl:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rtd:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	75.8	774	AB076888	AB076888 Homo sapi
2	25	75.8	1249	AX430285	AX430285 Sequence
3	25	75.8	2827	AK096600	AK096600 Homo sapi
4	25	75.8	3061	AX430418	AX430418 Sequence
5	25	75.8	3391	BC030660	BC030660 Homo sapi
6	25	75.8	177540	AC006538	AC006538 Homo sapi
7	24	72.7	597	AY056037	AY056037 Homo sapi
8	24	72.7	597	AY059641	AY059641 Homo sapi
9	21.8	66.1	1875	BC008065	BC008065 Homo sapi
10	21.8	66.1	2505	AB076889	AB076889 Homo sapi
11	21.8	66.1	4167	AX477382	AX477382 Sequence
12	21.8	66.1	68052	AC100691	AC100691 Mus muscu
13	21.8	66.1	87845	AL583860	AL583860 Human DNA
14	21.8	66.1	143299	AL353619	AL353619 Human DNA
15	21.8	66.1	143857	AC021583	AC021583 Homo sapi
16	21.8	66.1	164500	AC118945	AC118945 Rattus no
17	21.8	66.1	199612	AC021582	AC021582 Homo sapi
18	21.8	66.1	226889	AC094343	AC094343 Rattus no
19	21.4	64.8	181375	AC118369	AC118369 Rattus no
20	20.8	63.0	65522	AC115691	AC115691 Mus muscu
21	20.8	63.0	129121	AC120343	AC120343 Mus muscu
22	20.8	63.0	331657	AC125157	AC125157 Mus muscu
23	20.6	62.4	163118	AC090973	AC090973 Oryza sat
24	20.6	62.4	168372	AP002746	AP002746 Oryza sat
25	20.6	62.4	179714	AP002743	AP002743 Oryza sat
26	20.4	61.8	51661	AC116891	AC116891 Mus muscu
27	20.4	61.8	234945	AC107805	AC107805 Mus muscu
28	20.2	61.2	1442	AF356877	AF356877 Homo sapi
29	20.2	61.2	1452	BC000408	BC000408 Homo sapi
30	20.2	61.2	1490	S70154	S70154 cytosolic a
31	20.2	61.2	4412	AB062937	AB062937 Macaca fa
32	20.2	61.2	8430	F210832S05	F210832S05 Homo sapi
33	20.2	61.2	60942	AC124611	AC124611 Mus muscu
34	20.2	61.2	106794	AC114212	AC114212 Rattus no
35	20.2	61.2	126901	AC091518	AC091518 Mus muscu
36	20.2	61.2	138610	AC110097	AC110097 Rattus no
37	20.2	61.2	154076	AC023273	AC023273 Homo sapi
38	20.2	61.2	165802	AL135914	AL135914 Human DNA
39	20.2	61.2	194335	AC102586	AC102586 Mus muscu
40	20.2	61.2	201320	AC073816	AC073816 Mus muscu
41	20	60.6	94386	HS452M16	AL024493 Human DNA
42	20	60.6	109770	AC073763_3	Continuation (4 of
43	20	60.6	138428	AC124911	AC124911 Takifugu
44	20	60.6	186243	AC073740	AC073740 Mus muscu
45	20	60.6	230760	AC073714	AC073714 Mus muscu

## ALIGNMENTS

RESULT 1  
AB076888 774 bp mRNA PRI 25-JUN-2002  
LOCUS Homo sapiens mRNA for Di-Ras1, complete cds.  
DEFINITION AB076888  
ACCESSION AB076888  
VERSION AB076888.1 GI:21624247  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens  
brain cDNA.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Kontani,K., Tada,M., Ogawa,T., Okai,T. and Katada,T.  
TITLE Di-Ras: A Distinct Subgroup of Ras-family GTPases with Unique

Pred. No. is the number of results predicted by chance to have a



REFERENCE 1  
 AUTHORS Chalup, M.S., Altus, C.M., Lincoln, S.E., Dufour, G.E. and Jackson, S.  
 TITLE Molecules for disease detection and treatment  
 JOURNAL Patent: WO 0240715-A 200 23-MAY-2002;  
 INCYTE GENOMICS INC (US)

FEATURES  
 source 1. .3061  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /note="Incyte ID NO: LI:332161.1:2000SEP08"

BASE COUNT 606 a 840 c 927 g 688 t  
 ORIGIN

Query Match 75.8%; Score 25; DB 6; Length 3061;  
 Best Local Similarity 84.8%; Pred. No. 1.1;  
 Matches 28; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 1 GGGGATCCATCCGGACAGATACGATTAC 33  
 b 42 GCGGGAGAGATGCCGGACAGATACGATTAC 74

RESULT 5  
 BC030660 3391 bp mRNA linear PRI 21-MAY-2002  
 LOCUS  
 DEFINITION Homo sapiens, similar to R1g protein, clone MGC:33391  
 IMAGE:4814337, mRNA, complete cds.

ACCESSION BC030660.1 GI:21040534  
 VERSION  
 KEYWORDS MGC.  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 3391)  
 AUTHORS Strausberg, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (20-MAY-2002) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 CONTACT: MGC help desk  
 Email: [cgabbs-rt@mail.nih.gov](mailto:cgabbs-rt@mail.nih.gov)  
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
 CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki  
 Toshiyuki and Piero Carninci (RIKEN)  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Institute for Systems Biology  
<http://www.systemsbio.org>  
 contact: amadan@systemsbiology.org  
 Anup Madan, Jessica Fahney, Erin Helton, Mark Ketteman, Anuradha  
 Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whitting

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
 Series: IRAX Plate: 46 Row: e Column: 8  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: Genomescan gene  
 prediction.

FEATURES  
 source 1. .3391  
 Location/Qualifiers  
 1. .3391  
 /organism="Homo sapiens"  
 /db\_xref="LocusID:148252"  
 /db\_xref="taxon:9606"  
 /clone="MGC:33391 IMAGE:4814337"  
 /tissue\_type="Brain, hippocampus"  
 /clone\_lib="NIH MGC\_95"  
 /lab\_host="DH10B"  
 /note="Vector: pBluescript"  
 138. .734  
 /codon\_start=1  
 CDS

/product="similar to R1g protein"  
 /protein\_id="AAH30660.1"  
 /db\_xref="GI:21040534"  
 /translation="MEQSDNDRVAVVFGAGVGSSTVLFYKGRPDYPTTIDTY  
 KOVIGCDKSVCTLOITDTGSHQFPAVMDLSTSKGAFITVSVYSKQSELEGIYK  
 LTVQIKGSVEDIPVMLVGNKCDQREVDYRRAQAAQDMKCAFMTSAKNYNVKEL  
 FOELLTLEFRRMNSLINDGRSKGRKRDVRAVKGCTLM"

BASE COUNT 650 a 996 c 1016 g 729 t  
 ORIGIN

Query Match 75.8%; Score 25; DB 9; Length 3391;  
 Best Local Similarity 84.8%; Pred. No. 1.1;  
 Matches 28; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 1 GGGGATCCATCCGGACAGATACGATTAC 33  
 Db 129 GCGGGAGAGATGCCGGACAGATACGATTAC 161

RESULT 6  
 AC006538/c 177540 bp DNA linear PRI 07-FEB-1999  
 LOCUS  
 DEFINITION Homo sapiens chromosome 19, BAC 41195 (CIT-B-31c16), complete  
 sequence.  
 ACCESSION AC006538  
 VERSION AC006538.1 GI:4235145  
 KEYWORDS HTG.  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 177540)  
 AUTHORS Lamerdin, J.E., McCready, P.M., Skowronski, E., Wiseman, V.,  
 Burkhardt-Schultz, K., Gordon, L., Dias, J., Ramirez, M., Stllwagen, S.,  
 Phan, H., Velasco, N., Do, L., Regala, M., Terry, A., Barnes, J.,  
 Dangnanan, L., Erler, A., Christensen, M., Georgescu, A., Avila, J.,  
 Liu, S., Altix, C., Andreise, T., Trankheim, M., Amico-Keller, G.,  
 Coefficient, J., Duarte, S., Lucas, S., Bruce, R., Thomas, P., Quan, G.,  
 Krommiller, B., Arellano, A., Sanders, C., Ow, D., Nolan, M., Trong, S.,  
 Kobayashi, A., Olsen, A.S. and Carrino, A.V.  
 Sequence analysis of a 3.5 kb contig in human 19p13.3 containing a  
 serine protease gene cluster

TITLE  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 177540)  
 AUTHORS Lamerdin, J.E.  
 TITLE Direct Submission  
 JOURNAL Submitted (07-FEB-1999) Joint Genome Institute, Lawrence Livermore  
 National Laboratory, 7000 East Ave., Livermore, CA 94551, USA  
 Map and sequence oriented from p telomere to centromere. BC41195  
 (CIT-B-31c16) is currently separated from cosmid R32203 (AC006275)  
 to the left by a sequence gap of approximately 6 kb, and overlaps  
 BAC 102889 (CIT-B-191n6, AC006130) to the right from bases 175,545  
 to 177,540. Additional map and sequence information are available  
 at: <http://www.bio.lnl.gov/db/seq/genome/genome.html>.

FEATURES  
 source 1. .177540  
 Location/Qualifiers  
 1. .177540  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="19"  
 /map="19p13.3 between CDC34 and D19S342"  
 /clone="CIT-B-31c16"  
 /clone\_lib="Cal tech CIT-B BAC library"  
 /note="LNL clone name: BC41195"  
 complement(1..144)  
 repeat\_region  
 /rpt\_family="LIMC/D"  
 283. .381  
 /rpt\_family="AluJo"  
 complement(399..437)  
 repeat\_region  
 /rpt\_family="(CAAA)n"  
 complement(438..738)  
 repeat\_region  
 /rpt\_family="AluSx"  
 complement(868..931)  
 repeat\_region  
 /rpt\_family="LIMB7"

misc_feature	1175. .1365	/note="predicted exon, program: gfall2exons_human_1.3, frame: 2, quality: good, score: 57.000"
repeat_region	1608. .1904	
repeat_region	/rpt_family="AluSg"	
repeat_region	complement(2133. .2231)	
repeat_region	/rpt_family="MER81"	
repeat_region	2739. .2935	
repeat_region	/rpt_family="AluJo"	
repeat_region	2946. .3247	
repeat_region	/rpt_family="AluSx"	
repeat_region	3256. .3384	
repeat_region	/rpt_family="L1MC3"	
repeat_region	3389. .3604	
repeat_region	/rpt_family="AluJo"	
repeat_region	3631. .3716	
repeat_region	/rpt_family="L1MD3"	
repeat_region	complement(4087. .4242)	
repeat_region	/rpt_family="AluJo"	
repeat_region	complement(4294. .4436)	
repeat_region	/rpt_family="AluJo"	
repeat_region	complement(4598. .4887)	
repeat_region	/rpt_family="AluSg"	
repeat_region	complement(4906. .5373)	
repeat_region	/rpt_family="L1MB7"	
repeat_region	complement(5382. .5669)	
repeat_region	/rpt_family="AluSx"	
repeat_region	complement(6425. .6609)	
repeat_region	/rpt_family="(GGCA)n"	
repeat_region	complement(6610. .6765)	
repeat_region	/rpt_family="(GAAA)n"	
repeat_region	complement(6766. .7065)	
repeat_region	/rpt_family="AluSx"	
repeat_region	7093. .7392	
repeat_region	/rpt_family="AluSx"	
repeat_region	7423. .7511	
repeat_region	/rpt_family="L1MA7"	
repeat_region	7539. .7839	
repeat_region	/rpt_family="AluSx"	
repeat_region	7852. .7938	
repeat_region	/rpt_family="L1MA7"	
repeat_region	8001. .8312	
repeat_region	/rpt_family="AluJo"	
repeat_region	complement(8771. .8897)	
repeat_region	/rpt_family="MER20"	
repeat_region	complement(9532. .9722)	
repeat_region	/rpt_family="MER20"	
repeat_region	complement(9970. .10143)	
repeat_region	/rpt_family="MER20"	
repeat_region	10144. .10405	
repeat_region	/rpt_family="AluJo"	
repeat_region	10421. .10630	
repeat_region	/rpt_family="L1M3"	
repeat_region	10631. .10756	
repeat_region	/rpt_family="AluSg"	
repeat_region	10757. .11058	
repeat_region	/rpt_family="AluY"	
repeat_region	11068. .11243	
repeat_region	/rpt_family="AluSg"	
repeat_region	11247. .11520	
repeat_region	/rpt_family="L1M3"	
repeat_region	11522. .11816	
repeat_region	/rpt_family="AluSg"	
repeat_region	complement(11827. .12033)	
repeat_region	/rpt_family="MER58A"	
repeat_region	12255. .12555	
repeat_region	/rpt_family="AluJo"	
repeat_region	complement(12562. .12617)	
repeat_region	/rpt_family="(TA)n"	
repeat_region	12769. .12894	
repeat_region	/rpt_family="FLAM_A"	
repeat_region	complement(12901. .12951)	
repeat_region	/rpt_family="L1MC3"	
repeat_region	12960. .13013	
repeat_region	/rpt_family="LTR13"	
repeat_region	13013. .13408	
repeat_region	/rpt_family="LTR13"	
repeat_region	complement(13409. .13457)	
repeat_region	/rpt_family="L1MD3"	
repeat_region	13651. .13908	
repeat_region	/rpt_family="L1MB8"	
repeat_region	13906. .14116	
repeat_region	/rpt_family="L1MC2"	
repeat_region	complement(14173. .14468)	
repeat_region	/rpt_family="AluY"	
repeat_region	14497. .14811	
repeat_region	/rpt_family="AluSx"	
repeat_region	14812. .15011	
repeat_region	/rpt_family="(GGCA)n"	
repeat_region	15689. .16007	
repeat_region	/rpt_family="AluJo"	
repeat_region	16008. .16128	
repeat_region	/rpt_family="MLT1F"	
repeat_region	16130. .16430	
repeat_region	/rpt_family="AluSp"	
repeat_region	16432. .16563	
repeat_region	/rpt_family="(GAAA)n"	
repeat_region	16575. .16891	
repeat_region	/rpt_family="AluSx"	
repeat_region	16892. .16930	
repeat_region	/rpt_family="(GCAA)n"	
repeat_region	16989. .17073	
repeat_region	/rpt_family="MLT1F"	
repeat_region	17178. .17479	
repeat_region	/rpt_family="AluSx"	
repeat_region	17481. .18206	
repeat_region	/rpt_family="MER21B"	
repeat_region	18240. .18329	
repeat_region	/rpt_family="MLT1F"	
repeat_region	18380. .18748	
repeat_region	/rpt_family="MSTD"	
repeat_region	complement(19030. .19229)	
repeat_region	/rpt_family="MER20"	
repeat_region	complement(19579. .19692)	
repeat_region	/note="predicted exon, program: gfall2exons_human_1.3, frame: 1, quality: good, score: 53.000"	
repeat_region	complement(19953. .20160)	
repeat_region	/rpt_family="MER20"	
repeat_region	complement(20388. .20585)	
repeat_region	/rpt_family="MER20"	
repeat_region	20846. .20994	
repeat_region	/rpt_family="MIR"	
repeat_region	complement(21068. .21332)	
repeat_region	/rpt_family="AluSx"	
repeat_region	complement(21372. .21671)	
repeat_region	/rpt_family="AluSg"	
repeat_region	complement(21681. .21811)	
repeat_region	/rpt_family="FLAM_C"	
repeat_region	21837. .22056	
repeat_region	/rpt_family="MER58A"	
repeat_region	22240. .22539	
repeat_region	/rpt_family="AluSx"	
repeat_region	22697. .22831	
repeat_region	/rpt_family="AluJo"	
repeat_region	22833. .23132	
repeat_region	/rpt_family="AluSx"	
repeat_region	23136. .23412	
repeat_region	/rpt_family="L1MC3"	
repeat_region	complement(24002. .24172)	
repeat_region	/rpt_family="MER20"	
repeat_region	complement(24191. .24313)	
repeat_region	/rpt_family="MIR"	
repeat_region	24410. .24707	
repeat_region	/rpt_family="AluSg"	
repeat_region	complement(24739. .24926)	
repeat_region	/rpt_family="MER58A"	

```
repeat_region 24928..25229
/rpt_family="Alusx"

Query Match 75.8%; Score 25; DB 9; Length 177540;
Best Local Similarity 84.8%; Pred. No. 0.99;
Matches 28; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 GGGGATCATCGCGAAGACAGTACGATTAC 33
Db 73015 GCGGGGAAGATCGCGAAGACAGTACGATTAC 72983

RESULT 7
AY056037
LOCUS Homo sapiens Rlg protein mRNA, complete cds. PRI 24-JUL-2002
ACCESSION AY056037
VERSION AY056037.1 GI:16555333
KEYWORDS Homo sapiens.
SOURCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 597)
AUTHORS Ellis,C.A., Vos,M.D., Howell,H., Vallecorsa,T., Fults,D.W. and
Clark,G.J.
TITLE Rlg is a novel Ras-related protein and potential neural tumor
suppressor
Proc. Natl. Acad. Sci. U.S.A. 99 (15), 9876-9881 (2002)
JOURNAL 12107278
PUBMED 2 (bases 1 to 597)
REFERENCE Castro,A.F. and Quilliam,L.A.
AUTHORS Identification of Rlg, a novel Ras family member sharing homology
TITLE with NOX2 and Rap
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 597)
AUTHORS Castro,A.F. and Quilliam,L.A.
TITLE Direct Submission
JOURNAL Submitted (13-SEP-2001) Biochemistry and Molecular Biology, Indiana
University School of Medicine, 635 Barnhill Drive, MS-4053,
Indianapolis, IN 46202, USA
FEATURES
source Location/Qualifiers
1..597
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19p13.3"
/tissue_type="brain; anaplastic oligodendroglioma"
/note="derived from dbEST AI497811"
1..597
/note="Ras family GTPase"
/codon_start=1
/product="Rlg protein"
/protein_id="AAL23715.1"
/db_xref="GI:16555334"
/translation="MPEOSNDYRVYVVGAGVGKSSLVLPVKGTFPDYIPTEDY
ROYISCDKSVCTLOITPTGSHOPPMQRISISKGAFILVFSYTSKQSELEGPYIK
LIVQIKSVEDIPVLYGNKCDTQREVDYREAOVAQEMKCAFMETSAKMNVKEL
FOELLTETRRNSLNDGKRSQKRTDVKGKCTLM"
BASE COUNT 134 a 178 c 195 g 90 t
ORIGIN

Query Match 72.7%; Score 24; DB 9; Length 597;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 ATGCCGGAACAGTACGATTAC 33
Db 1 ATGCCGGAACAGTACGATTAC 24

RESULT 8
AY059641
```

```
LOCUS AY059641 597 bp mRNA linear PRI 29-OCT-2001
DEFINITION Homo sapiens small GTP-binding tumor suppressor 1 mRNA, complete
cde
ACCESSION AY059641
VERSION AY059641.1 GI:16508175
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
REFERENCE 1 (bases 1 to 597)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE Molecular cloning of GBR1, a novel gene encoding a small
GTP-binding tumor suppressor
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 597)
AUTHORS Gong,L.
TITLE Direct Submission
JOURNAL Submitted (12-OCT-2001) Dept. of Cardiology, U.T. MD. Anderson
Cancer Center, 1515 Holcombe Blvd-449, Houston, TX 77030, USA
FEATURES
source Location/Qualifiers
1..597
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
1..597
/note="GBR1"
/codon_start=1
/product="small GTP-binding tumor suppressor 1"
/protein_id="AAL17968.1"
/db_xref="GI:16508176"
/translation="MPEOSNDYRVYVVGAGVGKSSLVLPVKGTFPDYIPTEDY
ROYISCDKSVCTLOITPTGSHOPPMQRISISKGAFILVFSYTSKQSELEGPYIK
LIVQIKSVEDIPVLYGNKCDTQREVDYREAOVAQEMKCAFMETSAKMNVKEL
FOELLTETRRNSLNDGKRSQKRTDVKGKCTLM"
BASE COUNT 134 a 178 c 195 g 90 t
ORIGIN

Query Match 72.7%; Score 24; DB 9; Length 597;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 ATGCCGGAACAGTACGATTAC 33
Db 1 ATGCCGGAACAGTACGATTAC 24

RESULT 9
BC008065
LOCUS BC008065 1875 bp mRNA linear PRI 12-JUL-2001
DEFINITION Homo sapiens, similar to CG8500 gene product, clone MGC:2364
IMAGE:3029902, mRNA, complete cds.
ACCESSION BC008065
VERSION BC008065.1 GI:14165524
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
REFERENCE 1 (bases 1 to 1875)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE Direct Submission
JOURNAL Submitted (21-MAY-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.ncl.nih.gov
COMMENT Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: DCTD/DRP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
```

http://www.systemsbioology.org  
contact: amadan@systemsbiology.org  
Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia  
Greene, Mark Kettelman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found  
through the T.M.A.G.E. Consortium/BLM at: <http://image.llnl.gov>  
Series: INAL Plate: 3 Row: p Column: 6.  
Location/Qualifiers

## FEATURES

source

1..1875  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="MGC:2364 IMAGE:3029902"  
/tissue\_type="Ovary, adenocarcinoma"  
/clone\_1lb="NH\_MGC\_9"  
/lab\_host="DH10B-R"  
/note="Vector: pOTB7"  
80..679  
/codon\_start=1  
/product="Similar to CG8500 gene product"  
/protein\_id="AAH08065.1"  
/db\_xref="gi:14165525"  
/translation="MPEOSNDYRAVAFGAGVGKSSLVLFVKGTFRESYIPTVEDTY  
ROVISCDSICTLOITDTGSHQFPAMORLSISKGHAPLVYSITSROSLEKPIYE  
QICEKGDVESPIPLVGNKCDSPRSREPOSSEARALARTMKCAPMETSAKLNHYKE  
LFQELNLEKRTVSLQIDGKSKQKREKLKGCVM"

## CDS

BASE COUNT 518 a 438 c 472 g 447 t  
ORIGIN

Query Match 66.1%; Score 21.8; DB 9; Length 1875;  
Best Local Similarity 92.0%; Pred No. 36;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

.QY 9 CATGCCGAGACAGATTAC 33  
||||| |||||||  
Db 79 CATGCTGACAGAGTACGATTAC 103

.RESULT 10  
AB076889 2505 bp mRNA linear PRI 25-JUN-2002  
LOCUS  
DEFINITION Homo sapiens mRNA for DI-Ras2, complete cds.  
ACCESSION AB076889  
VERSION AB076889.1 GI:21624249  
KEYWORDS  
SOURCE Homo sapiens adult brain cDNA to mRNA, clone\_1lb:lambda ZAPII human brain cDNA.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1  
AUTHORS Kontani,K., Tada,M., Ogawa,T., Okai,T. and Katada,T.  
TITLE DI-Ras: A Distinct Subgroup of Ras-family GTPases with Unique  
JOURNAL Biochemical Properties  
REFERENCE 2 (bases 1 to 2505)  
AUTHORS Kontani,K., Ogawa,T., Okai,T., Tada,M. and Katada,T.  
TITLE Direct Submission  
JOURNAL Submitted (24-DEC-2001) Toshiaki Katada, University of Tokyo,  
Department of Physiological Chemistry, Graduate School of  
Pharmaceutical Sciences, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033,  
Japan (E-mail:katadaemol.f.u-tokyo.ac.jp, Tel:81-3-5841-4750,  
Fax:81-3-5841-4751)

## FEATURES

source

1..2505  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="9"  
/map="9q21.33-22.33"  
/tissue\_type="brain"  
/clone\_1lb="lambda ZAPII human brain cDNA"  
/dev\_stage="adult"  
1..2505  
gene

## CDS

/gene="Di-ras2"  
31..630  
/gene="Di-ras2"  
/note="member of the Ras family  
small gtp-binding protein"  
/codon\_start=1  
/product="Di-Ras2"  
/protein\_id="BAC01116.1"  
/db\_xref="gi:21624250"  
/translation="MPEOSNDYRAVAFGAGVGKSSLVLFVKGTFRESYIPTVEDTY  
ROVISCDSICTLOITDTGSHQFPAMORLSISKGHAPLVYSITSROSLEKPIYE  
QICEKGDVESPIPLVGNKCDSPRSREPOSSEARALARTMKCAPMETSAKLNHYKE  
LFQELNLEKRTVSLQIDGKSKQKREKLKGCVM"

## BASE COUNT

622 a 611 c 630 g 642 t

## ORIGIN

Query Match 66.1%; Score 21.8; DB 9; Length 2505;  
Best Local Similarity 92.0%; Pred No. 36;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

.QY 9 CATGCCGAGACAGATTAC 33  
||||| |||||||  
Db 30 CATGCTGACAGAGTACGATTAC 54

.RESULT 11  
AX477382 4167 bp DNA linear PAT 12-AUG-2002  
LOCUS  
DEFINITION Sequence 36 from Patent WO0231152.  
ACCESSION AX477382  
VERSION AX477382.1 GI:22216612  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1  
AUTHORS Baughn,M.R., Ding,L., Elliott,V.S., Gandhi,A.R., Gietzen,K.J.,  
Griffin,J.A., Gururajan,R., Hatalla,A.J., Kearney,L., Khan,F.A.,  
Lal,P., Lee,E.A., Lu,D.A., Lu,Y., Nguyen,D.B., Arvizu,C.,  
Ramkumar,J., Tang,Y.T., Thangavelu,K., Thornton,M., Walla,N.K.,  
Warren,B.A., Xu,Y., Yao,M.G. and Yue,H.  
TITLE Intracellular signaling molecules  
JOURNAL Patent: WO 0231152-A 36 18-APR-2002;  
Incyte Genomics, Inc. (US)  
Location/Qualifiers

## FEATURES

source  
1..4167  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/note="Incyte ID No: 490576CB1"

BASE COUNT 1098 a 911 c 960 g 1198 t  
ORIGIN

Query Match 66.1%; Score 21.8; DB 6; Length 4167;  
Best Local Similarity 92.0%; Pred No. 35;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

.QY 9 CATGCCGAGACAGATTAC 33  
||||| |||||||  
Db 160 CATGCTGACAGAGTACGATTAC 184

## RESULT 12

AC100691

LOCUS 68052 bp DNA linear HTG 22-NOV-2001  
DEFINITION Mus musculus clone RP23-168I21, LOW-PASS SEQUENCE SAMPLING.  
ACCESSION AC100691  
VERSION AC100691.1 GI:17048057  
KEYWORDS HTG; HTGS; PHASE0.  
SOURCE Mus musculus.  
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Mus.



```

REFERENCE
1 (bases 1 to 68052)
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Mus musculus, clone RP23-168I21
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 68052)
Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhalter, B.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Chopel, Y., Collangelo, M., Collins, S., Collamore, A., Cook, A.,
Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gargyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heatford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K.,
Lamasz, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
Maclean, C., MacDonald, P., Major, J., Margulis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J.,
Menus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Plerre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Roselli, M., Roy, A., Santos, R., Schauer, S., Schuback, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zaloun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L15958
Center clone name: 168_I_21

-----
* NOTE: This record contains 83 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1
* 713 812: contig of 712 bp in length
* 813 1530: contig of 718 bp in length
* 1531 1630: gap of 100 bp
* 1631 2351: contig of 721 bp in length
* 2352 2451: gap of 100 bp
* 2452 3179: contig of 728 bp in length
* 3180 3279: gap of 100 bp
* 3280 3980: contig of 701 bp in length
* 3981 4080: gap of 100 bp
* 4081 4819: contig of 739 bp in length
* 4820 4919: gap of 100 bp
* 4920 5652: contig of 733 bp in length
* 5653 5752: gap of 100 bp
* 5753 6455: contig of 703 bp in length
* 6456 6555: gap of 100 bp
* 6556 7280: contig of 735 bp in length
* 7291 7390: gap of 100 bp
* 7391 8113: contig of 723 bp in length
* 8114 8213: gap of 100 bp

8214 8941: contig of 728 bp in length
8942 9041: gap of 100 bp
9042 9758: contig of 717 bp in length
9759 9858: gap of 100 bp
9859 10578: contig of 720 bp in length
10579 10678: gap of 100 bp
10679 11404: contig of 726 bp in length
11405 11504: gap of 100 bp
11505 12227: contig of 723 bp in length
12228 12327: gap of 100 bp
12328 13051: contig of 724 bp in length
13052 13151: gap of 100 bp
13152 13873: contig of 722 bp in length
13874 13973: gap of 100 bp
13974 14702: contig of 729 bp in length
14703 14802: gap of 100 bp
14803 15530: contig of 728 bp in length
15531 15630: gap of 100 bp
15631 16358: contig of 728 bp in length
16359 16458: gap of 100 bp
16459 17174: contig of 716 bp in length
17175 17274: gap of 100 bp
17275 17983: contig of 709 bp in length
17984 18083: gap of 100 bp
18084 18803: contig of 720 bp in length
18804 18903: gap of 100 bp
18904 19629: contig of 726 bp in length
19630 19729: gap of 100 bp
19730 20445: contig of 716 bp in length
20446 20545: gap of 100 bp
20546 21269: contig of 724 bp in length
21270 21369: gap of 100 bp
21370 22094: contig of 725 bp in length
22095 22194: gap of 100 bp
22195 22914: contig of 720 bp in length
22915 23014: gap of 100 bp
23015 23751: contig of 737 bp in length
23752 23851: gap of 100 bp
23852 24593: contig of 742 bp in length
24594 24693: gap of 100 bp
24694 25425: contig of 732 bp in length
25426 25525: gap of 100 bp
25526 26260: contig of 735 bp in length
26261 26360: gap of 100 bp
26361 27070: contig of 710 bp in length
27071 27170: gap of 100 bp
27171 27902: contig of 732 bp in length
27903 28002: gap of 100 bp
28003 28722: contig of 720 bp in length
28723 28822: gap of 100 bp
28823 29514: contig of 692 bp in length
29515 29614: gap of 100 bp
29615 30337: contig of 723 bp in length
30338 30437: gap of 100 bp
30438 31142: contig of 705 bp in length
31143 31242: gap of 100 bp
31243 31970: contig of 728 bp in length
31971 32070: gap of 100 bp
32071 32794: contig of 724 bp in length
32795 32894: gap of 100 bp
32895 33629: contig of 735 bp in length
33630 33729: gap of 100 bp
33730 34453: contig of 724 bp in length
34454 34553: gap of 100 bp
34554 35284: contig of 731 bp in length
35285 35384: gap of 100 bp
35385 36107: contig of 723 bp in length
36108 36207: gap of 100 bp
36208 36931: contig of 724 bp in length
36932 37031: gap of 100 bp
37032 37758: contig of 727 bp in length
37759 37858: gap of 100 bp
37859 38559: contig of 701 bp in length

```

```

* 38560 38659: gap of 100 bp
* 38660 39381: contig of 722 bp in length
* 39382 39481: gap of 100 bp
* 39482 40200: contig of 719 bp in length
* 40201 40300: gap of 100 bp
* 40301 41013: contig of 719 bp in length
* 41020 41119: gap of 100 bp
* 41120 41849: contig of 730 bp in length
* 41850 41949: gap of 100 bp
* 41950 42683: contig of 734 bp in length
* 42684 42783: gap of 100 bp
* 42784 43510: contig of 727 bp in length
* 43511 43610: gap of 100 bp
* 43611 44330: contig of 720 bp in length
* 44331 44430: gap of 100 bp
* 44431 45116: contig of 686 bp in length
* 45117 45216: gap of 100 bp
* 45217 45945: contig of 729 bp in length
* 45946 46045: gap of 100 bp
* 46046 46756: contig of 711 bp in length
* 46757 46856: gap of 100 bp
* 46857 47587: contig of 731 bp in length
* 47588 48410: contig of 723 bp in length
* 48411 48510: gap of 100 bp
* 48511 48236: contig of 726 bp in length
* 49237 49336: gap of 100 bp
* 49337 50016: contig of 680 bp in length
* 50017 50116: gap of 100 bp
* 50117 50814: contig of 698 bp in length
* 50815 50914: gap of 100 bp
* 50915 51639: contig of 725 bp in length
* 51640 51739: gap of 100 bp
* 51740 52463: contig of 724 bp in length
* 52464 52563: gap of 100 bp
* 52564 53280: contig of 717 bp in length
* 53281 53380: gap of 100 bp
* 53381 54093: contig of 713 bp in length
* 54094 54193: gap of 100 bp
* 54194 54915: contig of 722 bp in length
* 54916 55015: gap of 100 bp
* 55016 55720: contig of 705 bp in length
* 55721 55820: gap of 100 bp
* 55821 56534: contig of 714 bp in length
* 56535 56634: gap of 100 bp

```

Query Match 66.1%; Score 21.8; DB 2; Length 68052;  
 Best Local Similarity 92.0%; Pred. No. 32;  
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 9 CATGCCGACAGTAACGATTAC 33  
 Db 4532 CATGCCGACAGTAACGATTAC 4556

RESULT 13  
 AL583860 87845 bp DNA linear PRI 13-NOV-2001  
 LOCUS Human DNA sequence from clone Rp11-16P8 on chromosome 10, complete  
 DEFINITION sequence.  
 ACCESSION AL583860 AC012252  
 VERSION AL583860.7 GI:16944093  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 87845)  
 AUTHORS Brown, J.  
 TITLE Direct Submission  
 JOURNAL Submitted (09-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

## COMMENT

On Nov 15, 2001 this sequence version replaced g1:15591326.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest. The following  
 abbreviations are used to associate primary accession numbers given  
 in the feature table with their source databases: Em', EMBL; SW',  
 SWISSPROT; Tr', TREMBL; Wp', WORMPEP; Information on the WORMPEP  
 database can be found at

[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence  
 was generated from part of bacterial clone contigs of human  
 chromosome 10, constructed by the Sanger Centre Chromosome 10  
 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr10>

Rp11-16P8 is from the library RPCI-11.1 constructed by the group of  
 Pieter de Jong. For further details see  
<http://www.choil.org/bacpac/home.htm>

VECTOR: PBAC3.6

IMPORTANT: This sequence is not the entire insert of clone  
 Rp11-16P8. It may be shorter because we sequence overlapping  
 sections only once, except for a short overlap.

The true left end of clone Rp11-16P8 is at 1 in this sequence. The  
 true left end of clone Rp11-551P18 is at 85846 in this sequence.  
 The true right end of clone Rp11-422P15 is at 44164 in this  
 sequence.

## FEATURES

## source

1..87845  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="10"  
 /clone="Rp11-16P8"  
 /clone\_11b="RPCI-11.1"  
 BASE COUNT 26908 a 19283 c 18854 g 22800 t  
 ORIGIN

Query Match 66.1%; Score 21.8; DB 9; Length 87845;  
 Best Local Similarity 78.8%; Pred. No. 32;  
 Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 GGGGATCCATCCCGGACAGTAACGATTAC 33  
 Db 22892 GGGGATCCATCCCGGACAGTAACGATTAC 22924

RESULT 14  
 AL353619 143299 bp DNA linear PRI 10-OCT-2001  
 LOCUS Human DNA sequence from clone Rp11-563G12 on chromosome  
 DEFINITION 9q21.33-22.33, complete sequence.  
 ACCESSION AL353619  
 VERSION AL353619.9 GI:16073639  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 143299)

AUTHORS  
 TITLE  
 JOURNAL

COMMENT Direct Submission  
 Submitted (10-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
 requests: clonerequest@sanger.ac.uk  
 On Oct 11, 2001 this sequence version replaced g1:15028633.  
 During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; SW: SWISSPROT; Tr: TrEMBL; Wp: WormPEP; information on the WormPEP database can be found at

[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr9>  
RP11-563G12 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see

<http://www.chori.org/dacpac/home.htm>

VECTOR: pBACe3.6  
IMPORTANT: This sequence is not the entire insert of clone RP11-563G12 it may be shorter because we sequence overlapping sections only once, except for a short overlap.  
The true right end of clone RP11-563G12 is at 143299 in this sequence. The true left end of clone RP11-61N16 is at 51116 in this sequence. The true right end of clone RP11-555F9 is at 2000 in this sequence.

#### FEATURES

##### source

```
1..143299
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/map="q21.33-22.33"
/clone="RP11-563G12"
/clone_lib="RPCI-11.2"
/misc_feature
4147..4230
/note="Single clone region. Assembly confirmed by restriction digest data."
```

BASE COUNT 43622 a 29112 c 29198 g 41367 t  
ORIGIN

#### Query Match

Best Local Similarity 92.0%; Score 21.8; DB 9; Length 143299;

Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
9 CATGCCGACAGATACGATTAC 33
||||| || |||||||||
Db 55609 CATGCCGACAGATACGATTAC 55585
```

RESULT 15  
AC021583/c 143857 bp DNA linear prt 29-JUN-2002  
LOCUS AC021583 Homo sapiens chromosome 9, clone RP11-1026H23, complete sequence.  
ACCESSION AC021583 GI:21629417  
VERSION AC021583.11  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Mammalia: Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 143857)  
TITLE Birren,B., Nussbaum,C. and Lander,E.  
JOURNAL Homo sapiens chromosome 9, clone RP11-1026H23  
REFERENCE Unpublished  
2 (bases 1 to 143857)  
AUTHORS Birren,B., Linton,L., Nussbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,

#### TITLE

##### JOURNAL

##### REFERENCE

##### AUTHORS

Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,  
Choepey,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,  
Dearellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,  
Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,  
Gardyna,S., Grant,G., Hagos,B., Heeford,A., Horton,L.,  
Howard,J.C., Johnson,R., Jones,C., Kain,L., Karats,A., Klein,J.,  
Lander,E., Lebecky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,  
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,  
McPheters,R., Meldrum,J., Menus,L., Morrow,J., Naylor,J.,  
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,  
Pierre,N., Pisenl,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,  
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,  
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,  
Zimmer,A. and Zody,M.

#### Direct Submission

Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 143857)

Birren,B., Linton,L., Nussbaum,C., Lander,E., Ali,A., Allen,N.,  
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,  
Boukhgalter,B., Brown,A., Camarata,J., Campiano,A., Chang,I.,  
Chazaro,B., Choepey,Y., Colangelo,M., Collins,S., Collymore,A.,  
Cooke,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
Faro,S., Ferreira,P., FitzGerald,M., Fitzhugh,W., Gage,D.,  
Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,  
Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,  
Johnson,R., Jones,C., Kamat,A., Karats,A., Kells,C., Labrecque,K.,  
Lamazares,R., Lander,E., Lebecky,J., Levine,R., Lindblad-Toh,K.,  
Liu,G., Maclean,C., Macdonald,P., Major,J., Marquis,N.,  
Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrum,J.,  
Menus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,  
Nicoli,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,  
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,  
Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,  
Rogov,P., Roman,J., Roselli,M., Roy,A., Santos,R., Schauer,S.,  
Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Testaye,S.,  
Theodore,J., Topham,K., Travers,M., Travis,N., Triggillo,J.,  
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
Young,G., Zalnoun,J., Zembek,L., Zimmer,A. and Zody,M.

#### Direct Submission

Submitted (21-JUN-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

4 (bases 1 to 143857)

Birren,B., Nussbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,  
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B.,  
Camarata,J., Chang,J., Chazaro,B., Choepey,Y., Collymore,A.,  
Cooke,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,  
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,  
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,  
Karats,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,  
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,  
McCarthy,M., Meldrum,J., Menus,L., Mihova,T., Mlenga,V.,  
Murphy,T., Naylor,J., Nguyen,C., Nicoli,R., Norbu,C., Norman,C.H.,  
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,  
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,  
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,  
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,  
Testaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zalnoun,J.,  
Zembek,L., Zimmer,A. and Zody,M.

#### Direct Submission

Submitted (29-JUN-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jun 29, 2002 this sequence version replaced: 1:21536044.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/kw/RepeatMasker.html>

#### Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR

Web site: <http://www.seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
 ----- Project Information  
 Center Project name: L5759  
 Center Clone name: 1026\_H23  
 -----  
 SP6 end overlaps Sanger clone AL3548843; T7 end overlaps Sanger  
 clone AL353619. We are only submitting the final 143.9 kilobases at  
 the T7 end.

## FEATURES

source location/Qualifiers  
 1. 143857  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="9"  
 /map="9"  
 /clone="RP11-1026H23"  
 /clone\_1lb="RPC1-11 Human Male BAC"  
 19. 278  
 /rpt\_family="HAL1b"  
 complement(451..789)  
 /rpt\_family="MLT1A1"  
 932..942  
 /note="<30 qual SNGL region"  
 959..965  
 /note="<30 qual SNGL region"  
 986..998  
 /note="<30 qual SNGL region"  
 1005..1009  
 /note="<30 qual SNGL region"  
 1059..1063  
 /note="<30 qual SNGL region"  
 1108..1114  
 /note="<30 qual SNGL region"  
 1133..1142  
 /note="<30 qual SNGL region"  
 1162..1167  
 /note="<30 qual SNGL region"  
 repeat\_region complement(2267..2426)  
 /rpt\_family="MER21B"  
 2484..2913  
 /rpt\_family="MER31B"  
 complement(2950..3122)  
 /rpt\_family="LTR33A"  
 3625..3663  
 /rpt\_family="AT\_r1ch"  
 4641..4683  
 /rpt\_family="AT\_r1ch"  
 complement(5259..5455)  
 /rpt\_family="MIR"  
 5537..5580  
 /rpt\_family="AT\_r1ch"  
 complement(6557..6711)  
 /rpt\_family="FRAM"  
 complement(7085..7256)  
 /rpt\_family="MIR"  
 7822..7978  
 /rpt\_family="MIR"  
 complement(8002..8427)  
 /rpt\_family="LTR56"  
 11612..11786  
 /rpt\_family="(GAAA)n"  
 11821..12214  
 /rpt\_family="MLT1J2"  
 complement(12574..13349)  
 /rpt\_family="L1PA16"  
 13356..14544  
 /rpt\_family="L1P"  
 complement(14548..14730)  
 /rpt\_family="L1MEC"  
 complement(14800..15265)  
 /rpt\_family="L1MEC"  
 repeat\_region complement(15337..15444)

repeat\_region /rpt\_family="L1MEC"  
 15719..15752  
 /rpt\_family="AT\_r1ch"  
 repeat\_region complement(15818..16111)  
 /rpt\_family="AluSq"  
 repeat\_region complement(16125..16404)  
 /rpt\_family="AluSc"  
 16972..16977  
 /note="<30 qual SNGL region"  
 16983..16991  
 /note="<30 qual SNGL region"  
 17075..17080  
 /note="<30 qual SNGL region"  
 complement(17596..18051)  
 /rpt\_family="L1M4"  
 repeat\_region complement(18086..18163)  
 /rpt\_family="L1MA10"  
 18177..18527  
 /rpt\_family="L1MDA"  
 18528..18671  
 /rpt\_family="L1MC/D"  
 20646..20674  
 /rpt\_family="AT\_r1ch"  
 repeat\_region complement(21399..21794)  
 /rpt\_family="MER11C"  
 repeat\_region complement(23176..23259)  
 /rpt\_family="MIR"  
 repeat\_region complement(23495..23612)  
 /rpt\_family="L1M4"  
 23613..23979  
 /rpt\_family="HEIC"  
 repeat\_region complement(23980..24091)  
 /rpt\_family="L1M4"  
 25843..26169  
 /rpt\_family="L1"  
 repeat\_region 26210..26378  
 /rpt\_family="L1"  
 repeat\_region 26548..27246

Query Match 66.1%; Score 21.8; DB 9; Length 143857;  
 Best Local Similarity 92.0%; Pred. No. 31;  
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Oy 9 CATGCCGAGACAGATACGATTAC 33  
 ||||| |||||||||  
 Db 99216 CATGCTGACGAGTACGATTAC 99192

Search completed: March 23, 2003, 16:00:35  
 Job time : 264.806 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 23, 2003, 13:56:37 ; Search time 38.371 Seconds  
(without alignments)  
1936.775 Million cell updates/sec

Title: US-09-873-546-2  
Perfect score: 33  
Sequence: 1 gggggatccatgcgcgcgaacagatcaacgattac 33

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: N.Geneseq\_101002.\*  
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
25: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	75.8	1249	24	ABO72525
2	25	75.8	3061	24	ABO72648
3	21.8	66.1	1087	22	AA160838
4	21.8	66.1	1108	22	AA159052
5	21.8	66.1	4167	24	AA037605
6	20.2	61.2	402	22	AA191981
7	20.2	61.2	459	21	AA002299
8	20.2	61.2	3249	21	AA078063
9	20.2	61.2	4011	23	AA594401

C	10	19.8	60.0	2187	22	AAE81507	White lupine secre
C	11	19.4	58.8	249487	24	ABN85733	Mouse genomic reg1
C	12	19.2	58.2	2081	21	AA078839	Human cancer assoc
C	13	19.2	58.2	5201	22	AAH90097	Human bone marrow
C	14	19.2	58.2	5294	22	AAH90071	Human bone marrow
C	15	19.2	58.2	5396	23	AA574120	DNA encoding novel
C	16	19.2	58.2	5404	24	AB162860	Breast cancer rela
C	17	19.2	58.2	5404	24	AB163081	Breast cancer rela
C	18	19.2	58.2	2944528	24	ABA03041	Listeria monocytog
C	19	19	57.6	93	20	AAH85454	Human artery perfu
C	20	19	57.6	93	20	AAH85455	Human artery perfu
C	21	19	57.6	439	22	AA524757	Human ovarian PCR-
C	22	19	57.6	546	22	AAH83390	Human ovarian tumo
C	23	19	57.6	585	22	AAH11047	Human CDNA clone (
C	24	19	57.6	1424	22	AAH15075	Human CDNA sequenc
C	25	19	57.6	3266	21	AA090055	Murine Th2-specific
C	26	18.6	56.4	354	24	ABN67790	Streptococcus poly
C	27	18.6	56.4	1230	23	AB112537	Drosophila melanog
C	28	18.6	56.4	2252	23	AB123126	Drosophila melanog
C	29	18.6	56.4	3230	23	AB112536	Drosophila melanog
C	30	18.4	55.8	956	14	AA050152	Serratia marcescen
C	31	18.4	55.8	957	20	AAH67274	Serratia spp. phos
C	32	18.4	55.8	957	21	AA256824	Serratia phospholi
C	33	18.4	55.8	1400	7	AAH60877	Sequence encoding
C	34	18.4	55.8	3001	21	AAH51782	Chromosome 13q31-q
C	35	18.4	55.8	3221	22	AAH14451	Human CDNA sequenc
C	36	18.2	55.2	1176	24	ABK63738	Rat sequence diffe
C	37	18.2	55.2	1716	24	ABO90086	M. capsulatus gene
C	38	18	54.5	564	22	AAH81660	Human differential
C	39	18	54.5	2623	23	AB110171	Drosophila melanog
C	40	18	54.5	3261	23	AB118857	Drosophila melanog
C	41	18	54.5	5326	23	AB110170	Drosophila melanog
C	42	18	54.5	6288	22	AA544629	Fission yeast full
C	43	18	54.5	16424	22	AAK68448	Human immune/haema
C	44	18	54.5	16424	22	AAK56177	Human immune/haema
C	45	18	54.5	39651	23	AB118856	Drosophila melanog

ALIGNMENTS

RESULT 1	
ABO72525	
ID ABO72525 standard; cDNA; 1249 BP.	
XX	
AC ABO72525:	
XX	
DT 03-SEP-2002 (first entry)	
XX	
DE Human MDDT encoding cDNA SEQ ID NO 77.	
XX	
KW Human: MDDT, disease detection and treatment molecule polynucleotide;	
KW proliferative disorder; hepatitis; cancer; AIDS;	
KW autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;	
KW rheumatoid arthritis; transgenic; gene therapy; antitumor; antitumor;	
KW hepatocytic; antitumor; antitumor; antitumor; antitumor; antitumor;	
KW antitumor; antitumor; antitumor; antitumor; antitumor; antitumor;	
XX	
OS Homo sapiens.	
XX	
PN WO200240715-A2.	
XX	
PD 23-MAY-2002.	
XX	
PF 06-SEP-2001; 2001WO-US27628.	
XX	
PR 06-SEP-2000; 2000US-230505P.	
PR 06-SEP-2000; 2000US-230514P.	
PR 06-SEP-2000; 2000US-230515P.	
PR 06-SEP-2000; 2000US-230517P.	
PR 06-SEP-2000; 2000US-230518P.	
PR 06-SEP-2000; 2000US-230519P.	

06-SEP-2000; 2000US-230595P.  
PR 06-SEP-2000; 2000US-230597P.  
PR 06-SEP-2000; 2000US-230598P.  
PR 06-SEP-2000; 2000US-230599P.  
PR 06-SEP-2000; 2000US-230610P.  
PR 06-SEP-2000; 2000US-230865P.  
PR 06-SEP-2000; 2000US-230988P.  
PR 06-SEP-2000; 2000US-230989P.  
PR 07-SEP-2000; 2000US-230951P.  
PR 07-SEP-2000; 2000US-231163P.  
PR 07-SEP-2000; 2000US-231167P.  
XX (INCYTE GENOMICS INC.  
PA Jackson S, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL;  
XX Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE, Dahl CR;  
PI Moniyama MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM;  
PI Gerstl EH, Peralta CH, David MH, Panzer SR, Flores V, Daffo A;  
PI Marwaha R, Chen AJ, Chang SC, Au AP, Imman RR;  
XX WPI: 2002-527544/56.  
XX P-PSDB: ABP51307.  
XX Novel human disease detection and treatment polypeptide, useful in  
PT diagnosis, prevention or treatment of cell proliferative disorders e.g.  
PT arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder  
PT e.g. AIDS -  
XX Claim 1; Page 350-351; 618pp; English.  
XX The invention relates to an isolated human disease detection and  
CC treatment (MDPT) polypeptide (I) selected from a polypeptide having a  
CC sequence selected from 254 sequences (ABP51231-ABP51484) given in the  
CC specification, a naturally occurring polypeptide comprising a sequence  
CC having at least 90% identity to (I) or a biologically active or  
CC immunogenic fragment of (I). (I) is useful for screening a compound for  
CC effectiveness as an agonist or antagonist, for screening a compound that  
CC specifically binds (I) or modulates the activity of (I), and for  
CC preparing a polyclonal or monoclonal antibody by hybridoma technology.  
CC Nucleic acids (II) (ABQ72449-ABQ72700) encoding (I) are useful for  
CC screening a compound for effectiveness in altering expression of a target  
CC polynucleotide comprising. Oligonucleotides and antihodies are useful for  
CC detecting MDP in a sample or for assessing toxicity of a test compound,  
CC in a diagnostic test for a condition or a disease associated with the  
CC expression of MDP in a biological sample, for detecting (I) in a sample,  
CC and for purifying (I) from a sample. A composition comprising (I), an  
CC agonist or antagonist is useful for treating a disease or condition  
CC associated with decreased or increased expression of functional MDP.  
CC (I) or (II) are useful for diagnosing, treating or preventing disorders  
CC associated with aberrant expression of MDP, where the disorders are  
CC selected from a cell proliferative disorder such as arteriosclerosis,  
CC cirrhosis, hepatitis, psoriasis, and cancer and an  
CC autoimmune/inflammatory disorder such as AIDS, Addison's disease,  
CC allergy, anaemia, asthma, atherosclerosis, gout, multiple sclerosis or  
CC rheumatoid arthritis. (II) are useful for creating knockin humanised  
CC animals or transgenic animals to model human diseases. In somatic or  
CC germ-line gene therapy, to generate a transcript image of a tissue or cell  
CC type, for detecting differences in the chromosomal location due to  
CC translocation or inversion among normal, carrier or affected individuals  
CC and as hybridisation probes for mapping naturally occurring genomic  
CC sequences.  
XX Sequence 1249 BP; 230 A; 436 C; 355 G; 228 T; 0 other;  
XX Query Match 75.8%; Score 25; DB 24; Length 1249;  
XX Best Local Similarity 84.8%; Pred. No. 0.16;  
XX Matches 28; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 1 GGGGATCATCGCGAGACAGATTAACATTAC 33  
Db 42 GCGGGAGATGCGCGAGACAGATTAACATTAC 74

RESULT 2  
ABQ72648  
ID ABQ72648 standard; cDNA; 3061 BP.  
XX  
AC ABQ72648;  
XX  
DT 03-SEP-2002 (first entry)  
XX  
DE Human MDP encoding cDNA SEQ ID NO 200.  
XX  
KW Human; MDP; disease detection and treatment molecule polynucleotide;  
KW proliferative disorder; hepatitis; psoriasis; cancer; AIDS;  
KW autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;  
KW rheumatoid arthritis; transgenic; gene therapy; antiatherosclerotic;  
KW hepatotropic; antiinflammatory; antipsoriatic; cytostatic; anti-HIV;  
KW antiallergic; antianaemic; antiasthmatic; antiatherosclerotic; antihypertensive; antineoplastic; antineuritic; gene; ss.  
XX Homo sapiens.  
XX  
OS  
XX  
PN WO200240715-A2.  
XX  
PD 23-MAY-2002.  
XX  
PF 06-SEP-2001; 2001WO-US27628.  
XX  
PR 06-SEP-2000; 2000US-230505P.  
PR 06-SEP-2000; 2000US-230514P.  
PR 06-SEP-2000; 2000US-230515P.  
PR 06-SEP-2000; 2000US-230517P.  
PR 06-SEP-2000; 2000US-230518P.  
PR 06-SEP-2000; 2000US-230519P.  
PR 06-SEP-2000; 2000US-230595P.  
PR 06-SEP-2000; 2000US-230597P.  
PR 06-SEP-2000; 2000US-230598P.  
PR 06-SEP-2000; 2000US-230599P.  
PR 06-SEP-2000; 2000US-230610P.  
PR 06-SEP-2000; 2000US-230865P.  
PR 06-SEP-2000; 2000US-230988P.  
PR 06-SEP-2000; 2000US-230989P.  
PR 07-SEP-2000; 2000US-230951P.  
PR 07-SEP-2000; 2000US-231163P.  
PR 07-SEP-2000; 2000US-231167P.  
XX  
PA (INCYTE GENOMICS INC.  
PI Jackson S, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL;  
PI Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE, Dahl CR;  
PI Moniyama MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM;  
PI Gerstl EH, Peralta CH, David MH, Panzer SR, Flores V, Daffo A;  
PI Marwaha R, Chen AJ, Chang SC, Au AP, Imman RR;  
XX WPI: 2002-527544/56.  
XX P-PSDB: ABP51431.  
XX Novel human disease detection and treatment polypeptide, useful in  
PT diagnosis, prevention or treatment of cell proliferative disorders e.g.  
PT arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder  
PT e.g. AIDS -  
XX Claim 1; Page 414; 618pp; English.  
XX The invention relates to an isolated human disease detection and  
CC treatment (MDPT) polypeptide (I) selected from a polypeptide having a  
CC sequence selected from 254 sequences (ABP51231-ABP51484) given in the  
CC specification, a naturally occurring polypeptide comprising a sequence  
CC having at least 90% identity to (I) or a biologically active or  
CC immunogenic fragment of (I). (I) is useful for screening a compound for  
CC effectiveness as an agonist or antagonist, for screening a compound that  
CC specifically binds (I) or modulates the activity of (I), and for  
CC preparing a polyclonal or monoclonal antibody by hybridoma technology.  
CC Nucleic acids (II) (ABQ72449-ABQ72700) encoding (I) are useful for  
CC screening a compound for effectiveness in altering expression of a target

CC polynucleotide comprising. Oligonucleotides and antibodies are useful for  
CC detecting MDDr in a sample or for assessing toxicity of a test compound,  
CC in a diagnostic test for a condition or a disease associated with the  
CC expression of MDDr in a biological sample, for detecting (I) in a sample,  
CC and for purifying (I) from a sample. A composition comprising (I), an  
CC agonist or antagonist is useful for treating a disease or condition  
CC associated with decreased or increased expression of functional MDDr.  
CC (I) or (II) are useful for diagnosing, treating or preventing disorders  
CC associated with aberrant expression of MDDr, where the disorders are  
CC selected from a cell proliferative disorder such as arteriosclerosis,  
CC cirrhosis, hepatitis, psoriasis, and cancer and an  
CC autoimmune/inflammatory disorder such as AIDS, Addison's disease,  
CC allergy, anaemia, asthma, atherosclerosis, gout, multiple sclerosis or  
CC rheumatoid arthritis. (II) are useful for creating knockin humanised  
CC animals or transgenic animals to model human diseases, in somatic or  
CC germ-line gene therapy, to generate a transcript image of a tissue or cell  
CC type, for detecting differences in the chromosomal location due to  
CC translocation or inversion among normal, carrier or affected individuals  
CC and as hybridisation probes for mapping naturally occurring genomic  
CC sequences.

CC Sequence 3061 BP; 606 A; 840 C; 927 G; 688 T; 0 other;

CC Query Match 75.8%; Score 25; DB 24; Length 3061;

CC Best Local Similarity 84.8%; Pred. No. 0.18;

CC Matches 28; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

CC QY 1 GGGGATCCATGCCGACAGACGATAC 33

CC Db 42 GCGGGAGATGCCGACAGACGATAC 74

CC RESULT 3

CC ID AAI60838 standard; cDNA; 1087 BP.

CC AC AAI60838;

CC DT 22-OCT-2001 (first entry)

CC DE Human polynucleotide SEQ ID NO 4827.

CC XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
CC KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokine; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia; ss.

CC KW Homo sapiens.

CC PN WO200153312-A1.

CC PD 26-JUL-2001.

CC PF 26-DEC-2000; 2000WO-US34263.

CC PR 21-JAN-2000; 2000US-0488725.

CC PR 25-APR-2000; 2000US-0552317.

CC PR 09-JUL-2000; 2000US-0598042.

CC PR 19-JUL-2000; 2000US-0620312.

CC PR 03-AUG-2000; 2000US-0653450.

CC PR 14-SEP-2000; 2000US-0662191.

CC PR 19-OCT-2000; 2000US-0693036.

CC PR 29-NOV-2000; 2000US-0727344.

CC PA (HYSE-) HYSEQ INC.

CC PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

CC PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;

CC PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

CC DR WPI, 2001-442253/47.

DR P-PSDB; AAM41682.

XX Novel nucleic acids and polypeptides, useful for treating disorders

PT such as central nervous system injuries -

XX Claim 1; SEQ ID NO 4827; 10078bp; English.

XX The invention relates to human nucleic acids (AAI57798-AAI61369) and  
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic.

CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide

CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and

CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic

CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: immune system suppression,

CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,

CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders.

CC Note: The sequence data for this patent did not form part of the printed  
CC specification.

CC Sequence 1087 BP; 270 A; 299 C; 321 G; 197 T; 0 other;

CC Query Match 66.1%; Score 21.8; DB 22; Length 1087;

CC Best Local Similarity 92.0%; Pred. No. 4.4;

CC Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

CC QY 9 CATGCCGACAGACGATAC 33

CC Db 150 CATGCCGACAGACGATAC 174

CC RESULT 4

CC ID AAI59052 standard; cDNA; 1108 BP.

CC AC AAI59052;

CC DT 22-OCT-2001 (first entry)

CC DE Human polynucleotide SEQ ID NO 1255.

CC XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
CC KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokine; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia; ss.

CC KW Homo sapiens.

CC PN WO200153312-A1.

CC PD 26-JUL-2001.

CC PF 26-DEC-2000; 2000WO-US34263.

CC PR 21-JAN-2000; 2000US-0488725.

CC PR 25-APR-2000; 2000US-0552317.

CC PR 09-JUL-2000; 2000US-0598042.

CC PR 19-JUL-2000; 2000US-0620312.

CC PR 03-AUG-2000; 2000US-0653450.

CC PR 14-SEP-2000; 2000US-0662191.

CC PR 19-OCT-2000; 2000US-0693036.

CC PR 29-NOV-2000; 2000US-0727344.

CC PA (HYSE-) HYSEQ INC.

CC PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

CC PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;

CC PI Wang Y, Zhang Y, Zhang J;

PI Zhao QA, Zhou P, Goodrich R, Dmanac RT;  
XX WPI: 2001-442253/47.  
DR P-PSDB; AAM39896.  
XX Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -  
XX  
PS Claim 1; SEQ ID NO 1255; 10078bp; English.  
XX  
CC The invention relates to human nucleic acids (AA15798-AA161369) and  
CC the encoded polypeptides (AAM38642-AA42213) with nootropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/Inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.  
XX  
SQ Sequence 1108 BP; 274 A; 306 C; 328 G; 200 T; 0 other;  
XX  
Query Match 66.1%; Score 21.8; DB 22; Length 1108;  
Best Local Similarity 92.0%; Pred. No. 4.4;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
XX  
-QY 9 CATGCCGACAGAGTACGATTAC 33  
DB 171 CATGCCGACAGAGTACGATTAC 195  
XXXXXXXXXXXXXXXXXXXXXXXXXXXX  
XXXXXXXXXXXXXXXXXXXXXXXXXXXX  
-RESULT 5  
AAD37605  
AAD37605 standard; cDNA; 4167 BP.  
XX  
AC AAD37605;  
XX  
DT 27-AUG-2002 (first entry)  
XX  
DE Human intracellular signalling (INTSIG-16) cDNA.  
XX  
XX Human; intracellular signalling; INTSIG-16; gene therapy; nausea; asthma;  
XX Crohn's disease; keratosis; cell proliferative disorder; Grave's disease;  
XX arteriosclerosis; hepatitis; leukaemia; autoimmune disorder; meningitis;  
XX acquired immunodeficiency syndrome; AIDS; allergy; dermatitis; psoriasis;  
XX Hashimoto's thyroiditis; neurological disorder; developmental disorder;  
XX gastrointestinal disorder; reproductive disorder; nootropic; antiemetic;  
XX Alzheimer's disease; dementia; Creutzfeldt-Jacob disease; dermatological;  
XX antiinflammatory; anaemia; nephrotropic; immunosuppressive; thyromimetic;  
XX cancer; neuroprotective; cerebroprotective; antibacterial; virucide;  
XX gynaecological; gene; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key  
FT Location/Qualifiers  
FT CDS  
FT 161..760  
FT /\*tag= a  
FT /product= "Human INSIG-16 protein"  
FT sig\_peptide 161..259  
FT /\*tag= b  
FT mat\_peptide 260..757  
FT /\*tag= c  
FT /product= "Mature INTSIG-16 protein"  
XX  
XX WO200231152-A2.

PD 18-APR-2002.  
XX  
XX 12-OCT-2001; 2001WO-US32090.  
PF  
XX  
PR 13-OCT-2000; 2000US-240871P.  
PR 30-OCT-2000; 2000US-244723P.  
PR 15-NOV-2000; 2000US-249402P.  
PR 22-NOV-2000; 2000US-252622P.  
PR 13-DEC-2000; 2000US-255622P.  
XX  
XX (INCYTE GENOMICS INC.  
XX  
XX Baughn MR, Ding L, Elliott VS, Gandhi AR, Gietzen KJ, Griffin JA;  
XX Gururajan R, Hatalla AJA, Kearney L, Khan FA, Lal P, Lee EA;  
XX Lu DAM, Lu Y, Nguyen DB, Arvizu C, Ramkumar J, Tang YT;  
XX Thangavelu K, Thornton M, Walla NK, Warren BA, Xu Y, Yao MG;  
XX Yue H;  
XX WPI: 2002-426285/45.  
XX P-PSDB; AAE23389.  
DR  
XX  
XX Novel intracellular signaling polypeptide useful for treating and  
XX preventing cell proliferative, autoimmune, neurological,  
XX gastrointestinal and reproductive and developmental disorders -  
XX  
XX Claim 91; Page 171-172; 177pp; English.  
XX  
XX  
XX The invention relates to an isolated intracellular signalling (INTSIG)  
XX polypeptide or a biologically active or immunogenic fragment of INTSIG.  
XX INTSIG is also useful for preparing a polyclonal or monoclonal antibody  
XX by hybridoma technology. INTSIG gene is useful in gene therapy. INTSIG  
XX is useful for treating or preventing disorders such as cell proliferative  
XX disorder e.g. keratosis, arteriosclerosis, hepatitis, psoriasis,  
XX leukaemia; autoimmune disorders e.g. AIDS, allergies, anaemia, asthma,  
XX Crohn's disease, dermatitis, Graves' disease, Hashimoto's thyroiditis;  
XX neurological disorders e.g. Alzheimer's disease, dementia, meningitis,  
XX Creutzfeldt-Jacob disease; gastrointestinal disorders e.g. nausea,  
XX reproductive disorders e.g. menstrual disorders, ectopic pregnancy and  
XX cancer of the breast and developmental disorders. The present sequence  
XX is human INTSIG-16 cDNA.  
XX  
SQ Sequence 4167 BP; 1098 A; 911 C; 960 G; 1198 T; 0 other;  
XX  
Query Match 66.1%; Score 21.8; DB 24; Length 4167;  
Best Local Similarity 92.0%; Pred. No. 5.3;  
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
XX  
-QY 9 CATGCCGACAGAGTACGATTAC 33  
DB 160 CATGCCGACAGAGTACGATTAC 184  
XXXXXXXXXXXXXXXXXXXXXXXXXXXX  
XXXXXXXXXXXXXXXXXXXXXXXXXXXX  
-RESULT 6  
AA191981/C  
AA191981 standard; cDNA; 402 BP.  
XX  
AC AA191981;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human polynucleotide SEQ ID NO 12041.  
XX  
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
XX tissue growth factor; immunomodulatory; cancer; leukaemia;  
XX nervous system disorders; arthritis; inflammation; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200164835-A2.  
XX  
XX 07-SEP-2001.  
XX



```

PF      26-FEB--2001; 2001MO-US04927.
XX
PR      28-FEB-2000; 2000US-0515126.
XX      18-MAY-2000; 2000US-0577409.
XX
PA      (HYSE-) HYSEQ INC.
XX
PI      Tang YF, Liu C, Drmanac RT;
XX
DR      WPI: 2001-514836/56.
XX      P-PSDB; AAOI2050.
XX
PT      Isolated nucleic acids and polypeptides, useful for preventing
PT      diagnosing and treating e.g. leukaemia, inflammation and immune
PT      disorders -
XX
PS      Claim 1; SEQ ID NO 12041; 1399pp + Sequence Listing; English.
XX
CC      The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC      the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC      cytokine, cell proliferation or cell differentiation or which may induce
CC      production of other cytokines in other cell populations. The
CC      polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC      peptide therapy. The polypeptides have various cytokine-like activities,
CC      e.g. stem cell growth factor activity, haematopoiesis regulating
CC      activity, tissue growth factor activity, immunomodulatory activity and
CC      activin/inhibin activity and may be useful in the diagnosis and/or
CC      treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC      inflammation.
CC      Note: The sequence data for this patent did not form part of the printed
CC      specification, but was obtained in electronic format directly from WIPO
CC      at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ      Sequence 402 BP; 88 A; 99 C; 127 G; 88 T; 0 other;
XX
Query Match          61.2%; Score 20.2; DB 22; Length 402;
Best Local Similarity 75.8%; Pred. No.20;
Matches    25; Conservative    0; Mismatches    8; Indels    0; Gaps    0
OY      1 GGGGATCCATGCCCGACAGATAACGATTAC 33
        || | ||||| ||||||||| || |
DB      331 GGCAGCTCATGCTGTGGACAGAGTAGCAATTTC 299

RESULT 7
AAC02299/c
ID      AAC02299 standard; CDNA; 459 BP.
XX
XX      AAC02299;
XX
DT      06-OCT--2000 (first entry)
XX
DE      Human secreted protein 5' EST, SEQ ID NO: 2297.
XX
KW      Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX      gene therapy; chromosome mapping; ss.
XX
OS      Homo sapiens.
XX
PN      EP1033401-A2.
XX
PD      06-SEP--2000.
XX
PF      21-FEB-2000; 2000EP-0200610.
XX
PR      26-FEB-1999; 99US-0122487.
XX
PA      (GEST ) GENSET.
XX
PI      Dumas Milne Edwards J, Duclert A, Giordano J;
XX
XX      WPI: 2000-500381/45.
XX      P-PSDB; AAG02293.
DR

```

xx	New nucleic acid that is a 5' expressed sequence tag (5' EST) for
pt	obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
pr	diagnostic, forensic, gene therapy and chromosome mapping procedures -
xx	
ps	Claim 1; SEQ ID 2297; 71pp + CD-ROM; English.
cc	The present sequence is one of a large number of 5' ESTs derived from
cc	mRNAs encoding secreted proteins. An ORF has been identified within the
cc	sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
cc	derived from 30 different tissues. EST sequences usually correspond
cc	mainly to the 3' untranslated region (UTR) of the mRNA because they are
cc	often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
cc	well suited for isolating cDNA sequences derived from the 5' ends of
cc	mRNAs and even in those cases where longer cDNA sequences have been
cc	obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
cc	mRNAs with intact 5' ends and can therefore be used to obtain full length
cc	cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
cc	gene therapy and chromosome mapping procedures. They are used to obtain
cc	upstream regulatory sequences and to design expression and secretion
cc	vectors.
xx	
sq	Sequence 459 BP; 102 A; 105 C; 145 G; 106 T; 1 other;
	Query Match 61.2%; Score 20.2; DB 21; Length 459;
	Best Local Similarity 75.8%; Pred. No. 21;
	Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
qy	1 GGGGATCCATCGCGGAACAGTAACGATTAC 33
db	325 GGCACTCCTCATCTGTGAACAGATGAGGCAATTC 293
	RESULT 8
	AAC78063
id	AAC78063 standard; cDNA; 3249 BP.
xx	
ac	AAC78063;
xx	
dt	08-FEB-2001 (first entry)
xx	
de	Human cancer associated gene sequence seq id NO:457.
xx	
kw	Human; cancer associated gene; cancer antigen; detection; cancer;
kw	dianosis; cytostatic; proliferative; vulnetary; immunomodulator;
kw	antidiabetic; antiaesthetic; antihemetic; antibacterial; antiviral;
kw	antiinflammatory; antithyroid; anti allergic; antipneumonia; cardiac;
kw	dermatologicall; neuroprotective; thrombolytic; coagulant; neotropic;
kw	vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
kw	immune disorder; haematopoietic cell disorder; autoimmune disorder;
kw	allergic reaction; graft versus host disease; organ rejection;
kw	haemostatic; thrombolytic; cardiovascular disorder; infection;
kw	neurological disease; drug screening; ss.
xx	
os	Homo sapiens.
xx	
pn	WO200055350-AI.
xx	
pd	21-SEP-2000.
xx	
pf	08-MAR-2000; 2000WO-USO5882.
xx	
pr	12-MAR-1999; 99US-0124270.
xx	
pa	(HUMA-) HUMAN GENOME SCT INC.
xx	
pi	Rosen CA, Ruben SM;
xx	
dr	WPI; 2000-587533/55.
xx	
xx	P-PSDB; AAB43854.
xx	
pt	Novel isolated nucleic acids comprising sequences encoding peptides
pt	useful for treating or diagnosing e.g. cancer -

XX Claim 1; Page 993-994; 2352zp; English.

CC AAC77607 to AAC78448 encode the human cancer associated proteins given  
CC In AAB43398 to AAB44239. The proteins can have activities based on the  
CC tissues and cells the genes are expressed in. Example of activities  
CC include: cytostatic; proliferative; vulnerary; immunomodulator;  
CC antiinflammatory; antisthmatic; antineumatic; antirheitic;  
CC antitumor; antithyroid; antihemorrhagic; antibacterial; antiviral;  
CC dermatological; neuroprotective; cardiant; thrombolytic; coagulant;  
CC neotropic; vasotropic; antipsoriatic and angiogenic. The  
CC polynucleotides and polypeptides can be used for preventing, treating or  
CC ameliorating medical conditions and diagnosing pathological conditions.  
CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from  
CC the present invention may be used to treat immune disorders by activating  
CC or inhibiting the proliferation, differentiation or mobilisation of  
CC immune cells, to treat disorders of haematopoietic cells, autoimmune  
CC rejection, allergic reactions, graft versus host disease and organ  
CC rejection, modulate haemostatic or thrombolytic activity, modulate  
CC inflammatory, cancers, cardiovascular disorders, neurological disease and  
CC bacterial or viral infections. The peptides, nucleotides, antibodies,  
CC agonists and antagonists may be also be used in drug screens. AAC78449 to  
CC AAC78457 and AAB44240 represent sequences used in the exemplification of  
CC the present invention.

XX Sequence 3249 BP; 900 A; 711 C; 748 G; 889 T; 1 other;

Query Match 61.2%; Score 20.2; DB 21; Length 3249;  
Best Local Similarity 75.8%; Pred. No. 27;  
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0

OY 1 GGGGATCCTATGCCGGACAGACTTACCATTC 33  
DB 3001 GGCACTCCTATCGTAGAAGAGTAGGAATTC 3033  
11 | 11111111111111111111111111111111

RESULT 9  
AAS94401  
ID AAS94401 standard; cDNA: 4011 BP.

XX AAS94401;  
XX 13-FEB-2002 (first entry)  
XX DE DNA encoding novel human diagnostic protein #30205.  
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX Homo sapiens.  
XX WO200175067-A2.  
XX PD 11-OCT-2001.  
XX PF 30-MAR-2001; 2001WO-US08631.  
XX PR 31-MAR-2000; 2000US-0540217.  
XX PR 23-AUG-2000; 2000US-0649167.  
XX PA (HYSE-) HYSEQ INC.  
XX DR Dmanac RT, Liu C, Tang YT;  
XX WIPI: 2001-639362/73.  
XX P-PSDB: ABG30214.

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XS Claim 1; SEQ ID NO 30205; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (III) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequencing sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPRO at ftp.wipro.int/pub/published\_pct\_sequences.

Sequence 4011 BP; 1075 A; 910 C; 953 G; 1073 T; 0 other;

Query Match 61.2%; Score 20.2; DB 23; Length 4011;  
Best Local Similarity 75.8%; Pred. No. 28;  
Matches 25; Conservative % 0; Mismatches 8; Indels 0; Gaps 0;

1 GGGGATCCATGCCCGACAGACTAACATTAC 33  
||| ||||||| ||||||||| |||  
Dn 3099 GGCACTCCTCATGTGACAGAGTAGGAATTTC 3131

RESULT 10  
AAAF81507/C  
ID AAFA81507 standard; DNA; 2187 BP.  
XX  
AC AAFA81507;  
XX  
DT 05-JUN-2001 (first entry)  
XX  
DE White lupine secretory acid phosphatase coding sequence.  
KM White lupine: secretory acid phosphatase; root; phosphor; ds.  
XX  
OS Lupinus albus.  
XX  
FH Location/Qualifiers  
FT 1..1917  
FT CDS /\*tag= a  
FT /product= "secretory acid phosphatase"  
XX  
PN JP2001029085-A.  
XX  
PD 06-FEB-2001.  
XX  
PF 07-JAN-2000; 2000JP-0001442.  
PR 18-MAY-1999; 99JP-0137273.  
XX  
PA (MITA ) MITSUI CHEM INC.  
XX  
DR WPI: 2001-224199/23.  
DR P-PSDB: AAB74496.  
XX  
PT Secretory acid phosphatase -  
XX  
PS Claim 1; Page 9-11; 17pp; Japanese.  
XX

The present invention provides the protein and coding sequences of a secretory acid phosphatase from the white lupine which combines to the surface of plant roots. The sequences are useful for producing plants

CC which can grow even under low phosphor condition.  
XX Sequence 2187 BP; 632 A; 447 C; 451 G; 657 T; 0 other;

Query Match 60.0%; Score 19.8; DB 22; Length 2187;  
Best Local Similarity 77.4%; Pred. No. 39;  
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 GGGGATCCATCCGAGACAGATACGATTA 32  
DB 1702 GGGGACTCATCCGAGATGAGAGACGATTA 1672

#### RESULT 11

ID ABN85733 standard; DNA: 249487 BP.

AC ABN85733;

DT 30-SEP-2002 (first entry)

XX Mouse genomic region containing the Itprc5 gene SEQ ID NO 3.

KM Mouse; Itprc5; taste; cell signalling; TC-ICS; food; pharmaceutical;  
KW taste cell-specific ion channel subunit; gene; ds.

XX Mus sp.

PN WO200254069-A1.

PD 11-JUL-2002.

PF 26-DEC-2001; 2001MO-US49808.

PR 29-DEC-2000; 2000US-259379P.

PR 21-DEC-2001; 2001US-0026188.

XX (REGC ) UNIV CALIFORNIA.

PI Zuker CS, Zhang Y;

DR WPI; 2002-583632/62.

XX Identifying modulators of taste signaling in taste cells for use in  
PT food and pharmaceutical industries to customize and regulate taste, by  
PT determining effect of the compound on a taste cell-specific ion channel  
PT subunit

XX Disclosure; Page 64-207; 306pp; English.

XX The invention relates to identifying (M1) a compound that modulates taste  
CC signalling in taste cells, by contacting the compound with a eukaryotic  
CC host cell or cell membrane which expresses a taste cell-specific ion  
CC channel subunit (TC-ICS), and determining a functional effect of the  
CC compound upon a transmembrane ion flux of a predetermined ion,  
CC identifying a compound that modulates taste signalling in taste cells.  
CC (M1) is useful for identifying a compound that modulates taste signalling  
CC in taste cells, for identifying a compound that binds to a taste cell  
CC specific ion channel subunit and for modulating taste signalling in taste  
CC cells of a mammal, in particular a human. Modulators identified by (M1)  
CC are used by the food and pharmaceutical industries to customize taste,  
CC e.g. as additives to food or medicine so that the food or medicine tastes  
CC different to the subject who ingests it. Bitter medicines can be made to  
CC taste less bitter and sweet substance can be enhanced. The modulators are  
CC useful for pharmacological and genetic modulation of taste signalling  
CC pathways. The taste modulators can be directly administered to mammalian  
CC subjects for modulation of taste in vivo. The present sequence is that of  
CC the mouse genomic region containing the Itprc5 gene of the invention.

XX Sequence 249487 BP; 64808 A; 61913 C; 61498 G; 61236 T; 32 other;

XX Query Match

XX Best Local Similarity 58.8%; Score 19.4; DB 24; Length 249487;  
XX 79.3%; Pred. No. 1.1e+02;

Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 GGGGATCCATCCGAGACAGATACGAT 30  
DB 198468 GTGGATCCATCCGAGACAGGAAATGAT 198496

#### RESULT 12

AC AC77839/C

ID AC77839 standard; cDNA: 2081 BP.

AC AC77839;

DT 08-FEB-2001 (first entry)

DE Human cancer associated gene sequence SEQ ID NO:233.

XX Human; cancer associated gene; cancer antigen; detection; cancer;  
KM diagnosis; cytostatic; proliferative; vulnerability; immunomodulator;  
KM antidiabetic; antidiabetic; antineoplastic; antitumor; antiviral;  
KM antineoplastic; antitumor; antidiabetic; antibacterial; cardiant;  
KM dermatological; neuroprotective; thrombolytic; coagulant; neotropic;  
KM vasotropic; antiproliferative; antineoplastic; gene therapy; inflammation;  
KM immune disorder; haematopoietic cell disorder; autoimmune disorder;  
KM allergic reaction; graft versus host disease; organ rejection;  
KM haemostatic; thrombolytic; cardiovascular disorder; infection;  
KM neurological disease; drug screening; ss.

XX Homo sapiens.

PN WO200055350-A1.

PD 21-SEP-2000.

PF 08-MAR-2000; 2000MO-US05882.

PR 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Ruben SM;

DR WPI; 2000-587533/55.

DR P-PSDB; AAB43630.

XX Novel isolated nucleic acids comprising sequences encoding peptides  
PT useful for treating or diagnosing e.g. cancer

XX Claim 1; Page 803-804; 2352pp; English.

XX AAC77607 to AAC78448 encode the human cancer associated proteins given  
CC in AAB43398 to AAB44239. The proteins can have activities based on the  
CC tissues and cells the genes are expressed in. Example of activities  
CC include: cytostatic; proliferative; vulnerability; immunomodulator;  
CC antidiabetic; antidiabetic; antineoplastic; antitumor; antiviral;  
CC antineoplastic; antitumor; antidiabetic; antibacterial; cardiant;  
CC dermatological; neuroprotective; cardiant; thrombolytic; coagulant;  
CC neotropic; vasotropic; antiproliferative and antineoplastic. The  
CC polynucleotides and polypeptides can be used for preventing, treating or  
CC ameliorating medical conditions and diagnosing pathological conditions.  
CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from  
CC the present invention may be used to treat immune disorders by activating  
CC or inhibiting the proliferation, differentiation or mobilisation of  
CC immune cells, to treat disorders of haematopoietic cells, autoimmune  
CC disorders, allergic reactions, graft versus host disease and organ  
CC rejection, modulate haemostatic or thrombolytic activity, modulate  
CC inflammation, cancers, cardiovascular disorders, neurological disease and  
CC bacterial or viral infections. The peptides, nucleotides, antibodies,  
CC agonists and antagonists may be also be used in drug screens. AAC78449 to  
CC AAC78457 and AAB44240 represent sequences used in the exemplification of  
CC the present invention.

XX Sequence 2081 BP; 594 A; 405 C; 400 G; 677 T; 5 other;



Query Match	58.2%;	Score 19.2;	DB 22;	Length 5294;
Best Local Similarity	75.0%;	Pred. No. 81;		
Matches 24; Conservative	0;	Mismatches 8;	Indels 0;	Gaps 0;

QY	2	GGGGATCCATGCCGGAACAGAGTAACGATTAC	33
Db	4563	GGAGATCCACGCGAGTAAGTAAGTCTGAC	45322

RESULT 15  
AAS74120/c

AC AAS74120;

DT 13-FEB-2002 (first entry)

DNA encoding novel human diagnostic protein #9924.

KM Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Llu C, Tang YT,

DR WPI; 2001-639362/73.

DR P-PSDB; A5G09933.

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutatio  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -

PS Claim 1; SEQ ID No 9924; 103pp; English

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. M5654197-NA934564 represent novel human diagnostic coding sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pct/sequences](http://wipo.int/pub/published_pct/sequences).

SQ Sequence 5396 BP; 1623 A; 1085 C; 1157 G; 1531 T; 0 other;

Query Match	58.2%;	Score 19.2;	DB 23;	Length 5396;
-------------	--------	-------------	--------	--------------

Best Local Similarity 75.0%; Pred. No. 81;  
Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0

QY	1	G G G G G A T C C A T G C C C G G A A C A G A G T A A C G A T T A	32
Db	5214	G G G G C A T C C A G G T G G G A A C A T C G T A A G G A C T A	5183

```
Search completed: March 23, 2003, 15:19:51
Job time : 113.871 secs
```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 23, 2003, 13:56:56 ; Search time 2058.86 Seconds

(without alignments)  
8438.824 Million cell updates/sec

Title: US-09-873-546-4

Perfect score: 597

Sequence: 1 atgcgcgaacagagtaacga.....gcaaatgcaccctcatgtga 597

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues 4109280

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenDbml:\*

1: gb\_da:\*

2: gb\_hcg:\*

3: gb\_in:\*

4: gb\_om:\*

5: gb\_ov:\*

6: gb\_pat:\*

7: gb\_ph:\*

8: gb\_pl:\*

9: gb\_pr:\*

10: gb\_ro:\*

11: gb\_sts:\*

12: gb\_sy:\*

13: gb\_un:\*

14: gb\_vl:\*

15: em\_ba:\*

16: em\_fun:\*

17: em\_hum:\*

18: em\_in:\*

19: em\_mu:\*

20: em\_om:\*

21: em\_or:\*

22: em\_ov:\*

23: em\_pat:\*

24: em\_ph:\*

25: em\_pl:\*

26: em\_ro:\*

27: em\_sts:\*

28: em\_un:\*

29: em\_vl:\*

30: em\_hcg\_hum:\*

31: em\_hcg\_inv:\*

32: em\_hcg\_other:\*

33: em\_hcg\_mus:\*

34: em\_hcg\_pln:\*

35: em\_hcg\_rod:\*

36: em\_hcg\_mam:\*

37: em\_hcg\_vrt:\*

38: em\_sy:\*

39: em\_hlgo\_hum:\*

40: em\_hlgo\_mus:\*

41: em\_hlgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	595.4	99.7	597	9	AY056037
2	595.4	99.7	597	9	AY056041
3	595.4	99.7	774	9	AB076888
4	595.4	99.7	2827	9	AK096600
5	595.4	99.7	3391	9	BC030660
6	595.4	99.7	177540	9	AC006538
7	583.4	97.7	1249	6	AX430295
8	582.8	97.6	3061	6	AX430418
9	501	83.9	597	10	AY059640
10	501	83.9	126901	2	AC091518
11	499.4	83.9	201320	2	AC073816
12	499.4	83.7	153386	2	AC103000
13	385.6	64.6	4412	9	AB062937
14	380.8	63.8	1875	9	BC008065
15	380.8	63.8	2505	9	AB076889
16	380.8	63.8	4167	6	AX477382
17	380.8	63.8	143299	9	AL353619
18	380.8	63.8	143857	9	AC021583
19	380.8	63.8	199612	9	AC021582
20	223.6	37.5	1971	9	AK096393
21	222	37.2	1286	6	AX077661
22	222	37.2	1495	9	HS096750
23	222	37.2	1511	9	BC005362
24	222	37.2	1561	9	AK021882
25	222	37.2	7242	9	AF202543
26	222	37.2	142396	9	AL157407
27	209.6	35.1	68052	2	AC100691
28	204	34.2	68052	2	AC100691
29	175.4	29.4	552	9	AF493915
30	175.4	29.4	552	9	HSRARP2B
31	175.4	29.4	1913	9	BC012362
32	175.4	29.4	140794	2	AC069415
33	175.4	29.4	165733	2	AC069072
34	175.4	29.4	198170	9	AC117394
35	175	29.3	2145	9	HSMB03027
36	174.2	29.2	240020	2	AC121587
37	174.2	29.2	289723	2	AC122935
38	167.4	28.0	552	10	AF386786
39	166.4	27.9	552	9	AF493914
40	166.4	27.9	555	9	HSRARP2
41	166.4	27.9	558	6	A08695
42	160.6	26.9	33443	2	AC014332
43	160.6	26.9	160817	3	AC008316
44	160.6	26.9	177028	3	AC008315
45	160.6	26.9	219579	3	AE003684

#### ALIGNMENTS

RESULT 1

AY056037 597 bp mRNA linear PRI 24-JUL-2002

LOCUS AY056037

DEFINITION Homo sapiens R1g protein mRNA, complete cds.

ACCESSION AY056037

VERSION AY056037.1 GI:16555333

KEYWORDS

SOURCE

ORGANISM Homo sapiens.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 597)

AUTHORS Ellis,C.A., Vos,M.D., Howell,H., Vallecorsa,T., Fults,D.W. and Clark,G.J.

TITLE R1g is a novel Ras-related protein and potential neural tumor

JOURNAL  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source

suppressor  
Proc. Natl. Acad. Sci. U.S.A. 99 (15), 9876-9881 (2002)  
12107278  
2 (bases 1 to 597)  
Castro, A.F. and Quilliam, L.A.  
Identification of R19, a novel Ras family member sharing homology  
with NOY2 and Rap  
Unpublished  
3 (bases 1 to 597)  
Castro, A.F. and Quilliam, L.A.  
Direct Submission  
Submitted (13-SEP-2001) Biochemistry and Molecular Biology, Indiana  
University School of Medicine, 635 Barnhill Drive, MS-4053,  
Indianapolis, IN 46202, USA  
Location/Qualifiers  
1..597  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="19"  
/map="19p13.3"  
/tissue\_type="brain; anaplastic oligodendroglioma"  
/note="derived from dbest A1497811"  
1..597  
/note="Ras family GTPase"  
/codon\_start=1  
/product="R19 protein"  
/protein\_id="AAL23715.1"  
/db\_xref="GI:1655334"  
/translation="MPEOSNDYRVVFGAGVGSLSVLRFKGFPRDYIPTIEDTY  
RQVTSCKDSVCTLOITDPTMGSHQFAPAMORLSISKHATILVSVTSKSLDELGIYK  
LIVQIKGSVEDIPVNLVGNKCDREVDREDAQVADBMKCAFETSAKKNYKEL  
FOELLTLEFRMSLNDIGKRSKOKRDRYKGTLM"

BASE COUNT 134 a 178 c 195 g 90 t

ORIGIN

Query Match 99.7%; Score 595.4; DB 9; Length 597;  
Best Local Similarity 99.8%; Pred. No. 5.9e-82;  
Matches 596; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCCGGAACAGAGTAACGATTACCGCGTGTGTTGCGGGCGGCGCGTGGCAAG 60  
DB 1 ATGCCGGAACAGAGTAACGATTACCGCGTGTGTTGCGGGCGGCGCGTGGCAAG 60  
QY 61 AGCTGCTGTGCTGCTGCTTCTGGAAGGACGCTTCCGGACACTACATCCCAACATC 120  
DB 61 AGCTGCTGTGCTGCTGCTTCTGGAAGGACGCTTCCGGACACTACATCCCAACATC 120  
QY 121 GAGGACACTACCGGAGGTGATCAGTGCAGCAAGAGGTGTCAGCTGCAGATCACA 180  
DB 121 GAGGACACTACCGGAGGTGATCAGTGCAGCAAGAGGTGTCAGCTGCAGATCACA 180  
QY 181 GACACACCGGAGCAGCAGCAGTTCGCCGCAATGACAGCGCTGTCTTCCAAGGGCCAC 240  
DB 181 GACACACCGGAGCAGCAGCAGTTCGCCGCAATGACAGCGCTGTCTTCCAAGGGCCAC 240  
QY 241 GCTTATCTGCTGTGCTTCTCCGTCACACCAAGCACTGCTGAGAGAGCTGGGGCCATC 300  
DB 241 GCTTATCTGCTGTGCTTCTCCGTCACACCAAGCACTGCTGAGAGAGCTGGGGCCATC 300  
QY 301 TCAACACTATGCTGATCAAGGAGGACGTCGAGAGATCCCTGATGCTCGTGGGC 360  
DB 301 TCAACACTATGCTGATCAAGGAGGACGTCGAGAGATCCCTGATGCTCGTGGGC 360  
QY 361 AACAACTGGATGAGACGAGCGGAGTGTGACACGCGGAGGCGGAGCGTGGCCCAA 420  
DB 361 AACAACTGGATGAGACGAGCGGAGTGTGACACGCGGAGGCGGAGCGTGGCCCAA 420  
QY 421 GAGTGAATGCGCTTTCATGAGACCTGCGCCAAAGATTAACCTACAAGCTCAAGAGCTC 480  
DB 421 GAGTGAATGCGCTTTCATGAGACCTGCGCCAAAGATTAACCTACAAGAGCTC 480  
QY 481 TTCCAGAGACTGCTGACGCTGAGAGCGCGGAAACATAGAGCTCAACATCAAGGCAAG 540  
DB 481 TTCCAGAGACTGCTGACGCTGAGAGCGCGGAAACATAGAGCTCAACATCAAGGCAAG 540

DB 481 TTCCAGAGACTGCTGACGCTGAGAGCGCGGAAACATAGAGCTCAACATCAAGGCAAG 540  
QY 541 CGCTCCGGAGAGAGAGAGACGACCGCGTCAAGGCAAAATGACCCCTCATGTGA 597  
DB 541 CGCTCCGGAGAGAGAGAGACGACCGCGTCAAGGCAAAATGACCCCTCATGTGA 597

RESULT 2  
LOCUS AY059641  
DEFINITION Homo sapiens small GTP-binding tumor suppressor 1 mRNA, complete cds.  
ACCESSION AY059641  
VERSION AY059641.1  
KEYWORDS GI:16508175  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
REFERENCE 1 (bases 1 to 597)  
AUTHORS Gong, L. and Wu, K.  
TITLE Molecular cloning of GBRSL, a novel gene encoding a small GTP-binding tumor suppressor  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 597)  
AUTHORS Gong, L.  
TITLE Direct Submission  
JOURNAL Submitted (12-OCT-2001) Dept. of Cardiology, U.T. MD. Anderson Cancer Center, 1515 Holcombe Blvd-449, Houston, TX 77030, USA  
Location/Qualifiers  
1..597  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="19"  
1..597  
/note="GBRSL"  
/codon\_start=1  
/product="small GTP-binding tumor suppressor 1"  
/protein\_id="AAL17968.1"  
/db\_xref="GI:16508176"  
/translation="MPEOSNDYRVVFGAGVGSLSVLRFKGFPRDYIPTIEDTY  
RQVTSCKDSVCTLOITDPTMGSHQFAPAMORLSISKHATILVSVTSKSLDELGIYK  
LIVQIKGSVEDIPVNLVGNKCDREVDREDAQVADBMKCAFETSAKKNYKEL  
FOELLTLEFRMSLNDIGKRSKOKRDRYKGTLM"

BASE COUNT 134 a 178 c 195 g 90 t

ORIGIN

Query Match 99.7%; Score 595.4; DB 9; Length 597;  
Best Local Similarity 99.8%; Pred. No. 5.9e-82;  
Matches 596; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCCGGAACAGAGTAACGATTACCGCGTGTGTTGCGGGCGGCGCGTGGCAAG 60  
DB 1 ATGCCGGAACAGAGTAACGATTACCGCGTGTGTTGCGGGCGGCGCGTGGCAAG 60  
QY 61 AGCTGCTGTGCTGCTGCTTCTGGAAGGACGCTTCCGGACACTACATCCCAACATC 120  
DB 61 AGCTGCTGTGCTGCTGCTTCTGGAAGGACGCTTCCGGACACTACATCCCAACATC 120  
QY 121 GAGGACACTACCGGAGGTGATCAGTGCAGCAAGAGGTGTCAGCTGCAGATCACA 180  
DB 121 GAGGACACTACCGGAGGTGATCAGTGCAGCAAGAGGTGTCAGCTGCAGATCACA 180  
QY 181 GACACACCGGAGCAGCAGCAGTTCGCCGCAATGACAGCGCTGTCTTCCAAGGGCCAC 240  
DB 181 GACACACCGGAGCAGCAGCAGTTCGCCGCAATGACAGCGCTGTCTTCCAAGGGCCAC 240  
QY 241 GCTTATCTGCTGTGCTTCTCCGTCACACCAAGCACTGCTGAGAGAGCTGGGGCCATC 300  
DB 241 GCTTATCTGCTGTGCTTCTCCGTCACACCAAGCACTGCTGAGAGAGCTGGGGCCATC 300  
QY 301 TCAACACTATGCTGATCAAGGAGGACGTCGAGAGATCCCTGATGCTCGTGGGC 360  
DB 301 TCAACACTATGCTGATCAAGGAGGACGTCGAGAGATCCCTGATGCTCGTGGGC 360



```

Db 301 TACAAGCTCATCGTGCAGATCAAGGCGACGCTGGAGGACATCCCGCTGATCTCTGTGGGC 360
QY 361 AACAGTGCATGAGACGACGCGGAGGTGAGACACGCGGAGGCGAGGCGTGGCCAA 420
Db 361 AACAGTGCATGAGACGACGCGGAGGTGAGACGCGGAGGCGAGGCGTGGCCAA 420
QY 421 GAGTGAAGTGCCTTTTCATGAGAGACTCGGCCAAGATGACTACAACTCAAGGAGACTC 480
Db 421 GAGTGAAGTGCCTTTTCATGAGAGACTCGGCCAAGATGACTACAACTCAAGGAGACTC 480
QY 481 TTCCAGAGAGCTGCTGACGCTGAGAGACGCGCGGAACATGAGCTCAACATCGAGCGCAAG 540
Db 481 TTCCAGAGAGCTGCTGACGCTGAGAGACGCGCGGAACATGAGCTCAACATCGAGCGCAAG 540
QY 541 CGCTCCGGAAGAGAGAGAGACGACGCGGTCAAGGGCAATGACACCTCATGTGA 597
Db 541 CGCTCCGGAAGAGAGAGAGAGACGACGCGGTCAAGGGCAATGACACCTCATGTGA 597

RESULT 3
LOCUS AB076888 774 bp mRNA linear PRI 25-JUN-2002
DEFINITION Homo sapiens mRNA for Di-Ras1, complete cds.
ACCESSION AB076888
VERSION AB076888.1 GI:21624247
KEYWORDS Homo sapiens adult brain cDNA to mRNA, clone_lib:lambda ZAPII human brain cDNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1
AUTHORS Kontani,K., Tada,M., Ogawa,T., Okai,T. and Katada,T.
TITLE Di-Ras: A Distinct Subgroup of Ras-family GTPases with Unique
JOURNAL Biochemical Properties
2 (bases 1 to 774)
AUTHORS Kontani,K., Ogawa,T., Okai,T., Tada,M. and Katada,T.
TITLE Direct Submission
JOURNAL Submitted (24-DEC-2001) Toshiaki Katada, University of Tokyo,
Department of Physiological Chemistry, Graduate School of
Pharmaceutical Sciences; 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033,
Japan (E-mail:katadaemol.f.u-tokyo.ac.jp, Tel:81-3-5841-4750,
Fax:81-3-5841-4751)
FEATURES
source location/Qualifiers
1..774
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19p13.3"
/tissue_type="brain"
/clone_lib="lambda ZAPII human brain cDNA"
/dev_stage="adult"
1..774
/gene="DI-Ras1"
/locus_tag="DI-Ras1"
/feature_id="1"
/feature_note="member of the Ras family
small GTP-binding protein"
/codon_start=1
/product="Di-Ras1"
/protein_id="BAC01115.1"
/db_xref="GI:21624248"
/translation="MPQSDNDYRVVYFAGAGVSKSLVLFYVGTFRDRIYPIPTDITY
ROYISCDKSVCTQITDITGSHOFPMKORLSISKAFILVESVTSQSLFEEIGPLYK
LIVQISGVSDIPLVMLVGNKCDTQREVPVTRQAAVAQDEWCAFMETSAKMNVVEL
FQELTLERNNSLINDKRSKQKRTREVRVKKCTLM"
BASE COUNT 154 a 268 c 239 g 113 t
ORIGIN
Query Match 99.7%; Score 595.4; DB 9; Length 774;
Best Local Similarity 99.8%; Pred. No. 5,6e-82;
Matches 596; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 ATGCCGGAACAGATACGATTACCGCGTGTGTGTTCTGGGGGCGGCGCTGGGCAAG 60
Db 100 ATGCCGGAACAGATACGATTACCGCGTGTGTGTTCTGGGGGCGGCGCTGGGCAAG 159
QY 61 AGCTGCTGTGTGTGCTGCTTCTGTGAAGGGACGTTCCGGACACATCAATCCACCATC 120
Db 160 AGCTGCTGTGTGTGCTGCTTCTGTGAAGGGACGTTCCGGACACATCAATCCACCATC 219
QY 121 GAGGACACCTACCGGCGAGGTGATGATGCTGGACAAAGGCTGTGACAGCTTCAGATACA 180
Db 220 GAGGACACCTACCGGCGAGGTGATGATGCTGGACAAAGGCTGTGACAGCTTCAGATACA 279
QY 181 GACACCAACCGGACACGACAGTTCGCCGATGAGGCGCTGTTCATCTCCAAAGGCGAC 240
Db 280 GACACCAACCGGACACGACAGTTCGCCGATGAGGCGCTGTTCATCTCCAAAGGCGAC 339
QY 241 GCCTTCATCTCTGTGTCTCCGTCACGACAGACGATGCTGAGAGAGCTGGGGCCATC 300
Db 340 GCCTTCATCTCTGTGTCTCCGTCACGACAGACGATGCTGAGAGAGCTGGGGCCATC 399
QY 301 TACAAGCTCATCTGTCAGATCAAGGCGCGTGGAGAGACATCCCGTATCTCTGTGGGC 360
Db 400 TACAAGCTCATCTGTCAGATCAAGGCGCGTGGAGAGACATCCCGTATCTCTGTGGGC 459
QY 361 AACAGTGCATGAGACGACGCGGAGGTGAGACACGCGGAGGCGAGCGGTGGCCAA 420
Db 460 AACAGTGCATGAGACGACGCGGAGGTGAGACACGCGGAGGCGAGCGGTGGCCAA 519
QY 421 GAGTGAAGTGCCTTTTCATGAGAGACTCGGCCAAGATGACTCAACATCGAGGAGCTC 480
Db 520 GAGTGAAGTGCCTTTTCATGAGAGACTCGGCCAAGATGACTCAACATCGAGGAGCTC 579
QY 481 TTCCAGAGAGCTGCTGACGCTGAGAGACGCGCGGAACATGAGCTCAACATCGAGCGCAAG 540
Db 580 TTCCAGAGAGCTGCTGACGCTGAGAGACGCGCGGAACATGAGCTCAACATCGAGCGCAAG 639
QY 541 CGCTCCGGAAGAGAGAGAGACGACGCGGTCAAGGGCAATGACACCTCATGTGA 597
Db 640 CGCTCCGGAAGAGAGAGAGACGACGCGGTCAAGGGCAATGACACCTCATGTGA 696

RESULT 4
LOCUS AK096600 2827 bp mRNA linear PRI 15-JUL-2002
DEFINITION Homo sapiens cDNA FLJ39281 fis, clone OCBF2011067, highly similar
to Homo sapiens cingulin mRNA.
ACCESSION AK096600
VERSION AK096600.1 GI:21756131
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens fetal brain cDNA to mRNA, clone_lib:OCBBF2
clone:OCBBF2011067.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1
AUTHORS Nishi,T., Oka,T., Nakagawa,S., Senoh,A., Mizuguchi,H., Inagaki,H.,
Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,
Yamamoto,J., Isono,Y., Kawai,Hio,T., Saito,K., Nishikawa,T.,
Yamamoto,J., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,
Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K.,
Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y.,
Sugano,S., Nagahari,K., Masubo,Y., Nagai,K. and Isogai,T.
NEDO human cDNA sequencing project
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
2 (bases 1 to 2827)
REFERENCE Isogai,T. and Yamamoto,J.
JOURNAL Direct Submission
Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0812 Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:

```

Research Association for Biotechnology (RAB): cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.): 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.

FEATURES  
source  
1.2827  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="OCBFF2011067"  
/tissue\_type="brain"  
/clone\_1ib="OCBFF2"  
/dev\_stage="fetal"  
/note="cloning vector: pME18SFL3"

BASE COUNT 688 a 750 c 983 g 406 t  
ORIGIN

Query Match 99.7%; Score 595.4; DB 9; Length 2827;  
Best Local Similarity 99.8%; Pred. No. 4.5e-82;

Matches 596; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

OY 1 ATGCCGGAACAGATACGATTACCGCGTGTGTTGGGGGCGGCGGTGGCAAG 60
    |||||||
DB 130 ATGCCGGAACAGATACGATTACCGCGTGTGTTGGGGGCGGCGGTGGCAAG 189
    |||||||
OY 61 ACCTCGCTGTGCTGCTTCCGTAAGGAGGACGTTCCGGACACACATCCCAACATC 120
    |||||||
DB 190 ACCTCGCTGTGCTGCTTCCGTAAGGAGGACGTTCCGGACACACATCCCAACATC 249
    |||||||
OY 121 GAGGACACTTACCGGAGGTGATACGCTCGACACAGAGGCTGTGACGCTGACATACA 180
    |||||||
DB 250 GAGGACACTTACCGGAGGTGATACGCTCGACACAGAGGCTGTGACGCTGACATACA 309
    |||||||
OY 181 GACACACCGGAGGAGGACAGTCCCGGCAATGCACAGCGCTGTCCAAAGGCCAC 240
    |||||||
DB 310 GACACACCGGAGGAGGACAGTCCCGGCAATGCACAGCGCTGTCCAAAGGCCAC 369
    |||||||
OY 241 GCCTTCATCTGTTGTTCCGTCACACAGACAGTCCGTAAGGAGGCTGGGGGCCATC 300
    |||||||
DB 370 GCCTTCATCTGTTGTTCCGTCACACAGACAGTCCGTAAGGAGGCTGGGGGCCATC 429
    |||||||
OY 301 TCAAGACTATGTCGATCAAGAGGCGACGCTGTGAGGACATCCCGTGTGCTGGGC 360
    |||||||
DB 430 TCAAGACTATGTCGATCAAGAGGCGACGCTGTGAGGACATCCCGTGTGCTGGGC 489
    |||||||
OY 361 AACAGTGCATGAGACGAGCGGAGGTGACACCGCGGAGGCGGAGCGGCGGCGCCAG 420
    |||||||
DB 490 AACAGTGCATGAGACGAGCGGAGGTGACACCGCGGAGGCGGAGCGGCGGCGCCAG 549
    |||||||
OY 421 GAGTGAAGTGGCTTTCATGAGACCTTCGCAAGATGATACAACTCAAGAGGACTC 480
    |||||||
DB 550 GAGTGAAGTGGCTTTCATGAGACCTTCGCAAGATGATACAACTCAAGAGGACTC 609
    |||||||
OY 481 TTCGAGAGCTGCTGACGCTGGAGACGCGCGAAACATGAGGCTCAACATCGACGGCAAG 540
    |||||||
DB 610 TTCGAGAGCTGCTGACGCTGGAGACGCGCGAAACATGAGGCTCAACATCGACGGCAAG 669
    |||||||
OY 541 GCCTCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 597
    |||||||
DB 670 GCCTCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 726
    |||||||

```

RESULT 5  
BC030660 3391 bp mRNA linear PRI 21-MAY-2002  
LOCUS  
DEFINITION Homo sapiens, similar to Rlg protein, clone MGC:33391  
IMAGE:4814337, mRNA, complete cds.  
ACCESSION BC030660  
VERSION BC030660.1 GI:21040534  
KEYWORDS MGC.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens

REFERENCE  
1 (bases 1 to 3391)  
AUTHORS  
TITLES  
JOURNAL  
Submitted (20-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 1A03, Bethesda, MD 20892-2590, USA

REMARK  
COMMENT  
NIH-MGC Project URL: <http://mgc.ncl.nih.gov>  
Contact: MGC help desk  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)

Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shireki Toshiyuki and Plero Carninci (RIKEN)  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Institute for Systems Biology  
<http://www.systemsbio.org>  
contact: [amadn@systemsbiology.org](mailto:amadn@systemsbiology.org)  
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
Series: IRAX Plate: 46 Row: e Column: 8  
This clone was selected for full length sequencing because it passed the following selection criteria: Genomescan gene prediction.

FEATURES  
source  
1.3391  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/db\_xref="locus:4814337"  
/clone="MGC:33391 IMAGE:4814337"  
/tissue\_type="Brain, hippocampus"  
/clone\_1ib="NIH MGC.95"  
/lab\_host="DH10B"  
/note="vector: pBluescript"

CDS  
138..734  
/codon\_start=1  
/product="similar to Rlg protein"  
/protein\_id="AAH30660.1"  
/db\_xref="GI:21040535"  
/translation="MPEOSNDYRVVVFAGGVGKSLVRFVKGFRDPTPTEDRY  
ROYISCDKSVCTLTQITDTTGGSHQPMARLISGHATILFVSLSKSLLELPYK  
LIVQIKSVEDIPVNLGNKCDTQREVDYDEAQAAGEMCAFMETSAKNNYVKEL  
FOELLTETLRNMSLINDGRSGOKRDRVVKCTLLM"

BASE COUNT 650 a 996 c 1016 g 729 t  
ORIGIN

Query Match 99.7%; Score 595.4; DB 9; Length 3391;  
Best Local Similarity 99.8%; Pred. No. 4.4e-82;  
Matches 596; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

OY 1 ATGCCGGAACAGATACGATTACCGCGTGTGTTGGGGGCGGCGGTGGCAAG 60
    |||||||
DB 138 ATGCCGGAACAGATACGATTACCGCGTGTGTTGGGGGCGGCGGTGGCAAG 197
    |||||||
OY 61 ACCTCGCTGTGCTGCTTCCGTAAGGAGGACGTTCCGGACACACATCCCAACATC 120
    |||||||
DB 198 ACCTCGCTGTGCTGCTTCCGTAAGGAGGACGTTCCGGACACACATCCCAACATC 257
    |||||||
OY 121 GAGGACACTTACCGGAGGTGATACGCTCGACACAGAGCGGTGTGACGCTCGACATACA 180
    |||||||
DB 258 GAGGACACTTACCGGAGGTGATACGCTCGACACAGAGCGGTGTGACGCTCGACATACA 317
    |||||||
OY 181 GACACACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 240
    |||||||
DB 318 GACACACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 377
    |||||||
OY 241 GCCTTCATCTGTTGTTCCGTCACACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300
    |||||||
DB 378 GCCTTCATCTGTTGTTCCGTCACACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 437
    |||||||

```

repeat_region	/rpt_family="AluSx" complement(868. .931) /rpt_family="L1MB7" 1175. .1365 /note="Predicted exon, program: grail2exons_human.1.3 frame: 2, quality: good, score: 57.000"
repeat_region	/rpt_family="AluSg" complement(2133. .2231) /rpt_family="MER81" 2739. .2935 /rpt_family="AluIo" 2945. .3247 /rpt_family="AluSx" 3256. .3384 /rpt_family="L1MC3" 3389. .3604 /rpt_family="AluIo" 3631. .3716 /rpt_family="L1MD3" complement(4087. .4242) /rpt_family="AluIo" complement(4294. .4436) /rpt_family="AluIo" complement(4598. .4687) /rpt_family="AluSg" complement(4906. .5373) /rpt_family="L1MB7" complement(5382. .5669) /rpt_family="AluSx" complement(6425. .6609) /rpt_family="GGGA)n" complement(6610. .6765) /rpt_family="GAAAn" complement(6766. .7065) /rpt_family="AluSx" 7093. .7392 /rpt_family="AluSx" 7423. .7511 /rpt_family="L1MA7" 7539. .7839 /rpt_family="AluSx" 7852. .7938 /rpt_family="L1MA7" 8001. .8312 /rpt_family="AluIo" complement(8711. .8897) /rpt_family="MER20" complement(9532. .9722) /rpt_family="MER20" complement(9970. .10143) /rpt_family="MER20" 10144. .10405 /rpt_family="AluIo" 10421. .10630 /rpt_family="L1M3" 10631. .10756 /rpt_family="AluSg" 10757. .11058 /rpt_family="AluY" 11066. .11243 /rpt_family="AluSg" 11247. .11520 /rpt_family="L1M3" 11522. .11816 /rpt_family="AluSg" complement(11827. .12033) /rpt_family="MER58A" 12255. .12555 /rpt_family="AluIo" complement(12562. .12617) /rpt_family="(TA)n" 12769. .12894

```
repeat_region /rpt_family="FLAM_A" complement(12901..12951)
repeat_region /rpt_family="L1MC3" 12960..13013
repeat_region /rpt_family="LTR13" 13013..13408
repeat_region /rpt_family="LTR13" complement(13409..13457)
repeat_region /rpt_family="L1MD3" 13651..13908
repeat_region /rpt_family="L1MB8" 13906..14116
repeat_region /rpt_family="L1MC2" complement(14117..14468)
repeat_region /rpt_family="ALUy" 14497..14811
repeat_region /rpt_family="ALUSx" 14812..15011
repeat_region /rpt_family="(GGGA)n" 15689..16007
repeat_region /rpt_family="ALUJo" 16008..16128
repeat_region /rpt_family="MLT1F" 16130..16430
repeat_region /rpt_family="ALUSp" 16432..16563
repeat_region /rpt_family="(GAAA)n" 16575..16891
repeat_region /rpt_family="ALUSx" 16892..16930
repeat_region /rpt_family="(GGAA)n" 16989..17073
repeat_region /rpt_family="MLT1F" 17178..17479
repeat_region /rpt_family="ALUSx" 17481..18306
repeat_region /rpt_family="MER21B" 18240..18329
repeat_region /rpt_family="MLT1F" 18380..18748
repeat_region /rpt_family="MSTD" complement(19030..19229)
repeat_region /rpt_family="MER20" complement(19579..19692)
misc_feature /note="Predicted exon, program: gtail2exons_human.1.3,
frame: 1, quality: good, score: 53.000"
repeat_region /rpt_family="MER20" complement(19953..20160)
repeat_region /rpt_family="MER20" complement(20388..20585)
repeat_region /rpt_family="MER20" 20846..20994
repeat_region /rpt_family="MIR" complement(21068..21332)
repeat_region /rpt_family="ALUSx" complement(21372..21671)
repeat_region /rpt_family="ALUSq" complement(21681..21811)
repeat_region /rpt_family="FLAM_C" 21837..22056
repeat_region /rpt_family="MER58A" 22240..22539
repeat_region /rpt_family="ALUSx" 22697..22831
repeat_region /rpt_family="ALUJo" 22833..23132
repeat_region /rpt_family="ALUSx" 23136..23412
repeat_region /rpt_family="L1MC3" complement(24002..24172)
repeat_region /rpt_family="MER20" complement(24191..24313)
repeat_region /rpt_family="MIR" 24410..24707
```

```
repeat_region /rpt_family="ALUSq" complement(24739..24926)
repeat_region /rpt_family="MER58A" 24928..25229
repeat_region /rpt_family="ALUSx"
```

Query Match 99.7% Score 595.4; DB 9; Length 177540;  
Best Local Similarity 99.8%; Pred. No. 2.2e-82;  
Matches 596; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 1 ATGCCGGAACAGATTAACGATTAACCGCGTGTGTGCGGGCGGCGCGTGGCCAG 60
DB 73006 ATCCCGAAGACAGATTAACGATTAACCGCGTGTGTGCGGGCGGCGCGTGGCCAG 72947
QY 61 AGCTCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 120
DB 72946 AGCTCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 72887
QY 121 GAGGACACCTACCGGAGGATGATCAGCTGCGACAAAGAGGTGTGTGTGTGTGTGT 180
DB 72886 GAGGACACCTACCGGAGGATGATCAGCTGCGACAAAGAGGTGTGTGTGTGTGT 72827
QY 181 GACACGACCGGAGGAGCCAGCAGTTCCGGCCATGACAGCGCTGTGTGTGTGTGT 240
DB 72826 GACACGACCGGAGGAGCCAGCAGTTCCGGCCATGACAGCGCTGTGTGTGTGT 72767
QY 241 GCGTTATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 300
DB 72766 GCGTTATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 72707
QY 301 TCAACGCTATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 360
DB 72706 TCAACGCTATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 72647
QY 361 AACAGTGGATGAGACGAGCGGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 420
DB 72646 AACAGTGGATGAGACGAGCGGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGT 72587
QY 421 GAGTGAAGTGGCTTTCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 480
DB 72586 GAGTGAAGTGGCTTTCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 72527
QY 481 TTCAGAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 540
DB 72526 TTCAGAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 72467
QY 541 CGCTCCGGGAACGAGAGGAGCAGACCGCGTGTGTGTGTGTGTGTGTGTGTGTGT 597
DB 72466 CGCTCCGGGAACGAGAGGAGCAGACCGCGTGTGTGTGTGTGTGTGTGTGTGT 72410
```

RESULT 7  
AX430295 1249 bp DNA linear PAT 28-JUN-2002  
LOCUS Sequence 77 from Patent WO0240715.  
DEFINITION AX430295  
ACCESSION AX430295  
VERSION AX430295.1 GI:21655659  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1  
AUTHORS Chalup,M.S., Altus,C.M., Lincoln,S.E., Dufour,G.E. and Jackson,S.  
TITLE Molecules for disease detection and treatment  
JOURNAL Patent: WO 0240715-A 77 23-MAY-2002;  
INCYTE GENOMICS INC (US)

FEATURES  
Source location/Qualifiers  
1..1249  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/note="Incyte ID No: LG:389870.1:2000SEP08"

BASE COUNT 230 a 436 c 355 g 228 t



```

/gene="Gbt1"
/note="Gbt1"
/codon_start=1
/product="small GTP-binding tumor suppressor 1"
/protein_id="A117967.1"
/db_xref="GI:16508174"
/translation="MPEOSNDYRVVVFAGAGVKSLSVLRVKGTPRDTYIPTEDYR
RVYKSCSVCTLIQITPTGSHOFPMARQSLISGHAFLVFSYTSQSDLSPIYK
LIVQKSGVEDIPIMLVGKNCDEQREGRVHIREAQAQEMKCAEMTSAKNNVKEI
FQELITETRSVSLSDYGRKSKQKADRIKCALM"

```

Query Match	83.98%	Score 501;	DB 10;	Length 597;
Best Local Similarity	89.98%	Pred. No. 1.7e-67;		
Matches 537; Conservative	0;	Mismatches 60;	Indels 0;	Gaps 0

QY	1	ATGCGGAAACAGATAAACGATATACCGGCTGGTGTTCGGGGGGCGCGCTGGGCAAG	60
Db	1	ATGCGAAGACAGACAAATGACTACCGTGTGTCTGTTCGGCGGAGCGGCTGGGCAAG	60
QY	61	AGCTCGCTGTGTCTGCGCTTCGTGAAGGGCACTGTCCGCACACCTCAATCCCAATC	120
Db	61	AGCTCGCTGTGTCTCGCTTTGTAAAGGGGACGTTTGTGACACCTACATCCCTACATA	120
QY	121	GAGGACACCTACCGGCAAGGTGATAGCTGCGACAAAGAGCGTGTGCACGTCGACANTACA	180
Db	121	GAGGACACCTACCGGCAAGGTGATAGCTGTGACAAAGAGCGTGTGCACACTCGACAGATCAG	180
QY	181	GACACCAACCGGACACCCACCAAGTTCGGGCAATGACAGGCGCTGTGCATCTCCAAAGGCGCAC	240
Db	181	GACACCAACCGGACACCCACCAAGTTCGGGCAATGACAGGCGCTGTGCATCTCCAAAGGCGCAC	240
QY	241	GCGCTTACTCTGTGTCTTCGTCACCAAGACAGTACGCTGAGAGAGCTGGGGGCCATC	300
Db	241	GCGCTTATCTCGTGTCTCTCGGTGACCAAGAGAGTCCGTGGATGTAGACCCCATC	300
QY	301	TACAAAGTCATCTCGGCAAGTCAAGGGAGCGGTGGAGACATCCCCGTGATGCTCTGGGCG	360
Db	301	TACAAAGTGTATCGGCAAGTCAAGGGAGCGGTGGAGACATCCCCATCACTGCTGTGGGG	360
QY	361	AACAAGTGCATGAGAGCGAGCGGAGAGTGGACACGGCGCAAGCGGCGTGGGCGCA	420
Db	361	AACAAGTGTATGAGAGCGAGCGGAGAGTGCACACGCGCAAGCGCGCTGGGCAAG	420
QY	421	GAGTGGAGTGTGCGCTTTCATGAGAGACCTCGGCCCAAGATGAATCAACGTCGAAGAGCTC	480
Db	421	GAGTGGAGTGTGCGCTTTCATGAGAAACCTCAGCAAGATGAATCAACGTCGAAGAGCTG	480
QY	481	TTCGAGAGGTGTGAGCGCTGGAGACGCGCGCGGAACATGAGCTCAACATCGACGCAAG	540
Db	481	TTCGAGAGGTGTGAGCGCTCGAGACGCGCGCGGAGGCTCAGCCTTCAGCTGTGACGCGCAAA	540
QY	541	CGCTCGGAGAGAGAGAGACAGACCGGCTCAAGGGCAAAATGCAACCCCAATGTGA	597
Db	541	CGTTCACGACGAGAGAGAGGCTGACACCCCATTAAGGGCAAGTGTCCGCTCATGTGA	597

RESULT 10	
AC091518	
LOCUS	AC091518 126901 bp DNA linear HTG 14-AUG-2002
DEFINITION	Mus musculus chromosome 10 clone mgs1-257j2 strain 129/SvJ, WORKING
DRAFT SEQUENCE,	5 unordered pieces.
AC091518	
AC091518.11	GI:22218471
HTG:	HTGS_PHASE1; HTGS_DRAFT.
SOURCE	house mouse.
ORGANISM	Mus musculus
Eukaryota;	Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia;	Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE	
AUTHORS	Kim,J., Shauli,S., Yao,Z., Andrews,G.K. and Roe,B.A.
TITLE	Mus musculus BAC Clone mgs1-257j2

JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 126901)  
AUTHORS Kim,J., Shaull,S., Yao,Z., Andrews,G.K. and Roe,B.A.  
TITLE Direct Submission  
JOURNAL Submitted (29-APR-2001) Department Of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
REFERENCE 3 (bases 1 to 126901)  
AUTHORS Kim,J., Shaull,S., Yao,Z., Andrews,G.K. and Roe,B.A.  
TITLE Direct Submission  
JOURNAL Submitted (14-AUG-2002) Department Of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA

COMMENT On Aug 14, 2002 this sequence version replaced gi:21955024.

Center: Department of Chemistry And Biochemistry  
The University Of Oklahoma  
Center code:UOKNOR

\* NOTE: This is a working draft sequence. It currently  
\* consists of 5 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

*	7755:	contig of 7755 bp in length
*	7756	7855: gap of unknown length
*	7856	29448: contig of 21593 bp in length
*	29449	29548: gap of unknown length
*	58825	58825: contig of 29277 bp in length
*	58826	588935: gap of unknown length
*	58919	89618: contig of 30693 bp in length
*	58926	gap of unknown length
*	89619	89718: gap of unknown length
*	89719	126901: contig of 37183 bp in length

FEATURES	Location/Qualifiers
source	1. .126901

```

/organism="Mus musculus"
/strain="129/SvJ"
/db_xref="taxon:10090"
/chromosome="10"
/clone="ngs1-257j2"
/clone_11b="Genome Systems mouse ES cell BAC libraries I
and 11"
BASE COUNT 30097 a 33386 c 32214 g 30700 t 504 others
ORIGIN

```

Query Match	83.9%;	Score 501;	DB 2;	Length 126901;
Best Local Similarity	89.9%;	Pred. No. 6.5e-68;		
Matches 537; Conservative	0;	Mismatches 60;	Indels 0;	Gaps 0;

OY	1	ATGCGGAAACAGAGTTAAACATTACCGCGTGGTGGTTCCGGGGGGGGCGGTGGGCAAC	60
Db	26982	ATGCGAAMACAGACAAATACATACCTGTGGTGGTGGTGGCGAGGCGGGGTGGGCAAG	27043
OY	61	AGCTCGCTGGTCTGCGCTTCTGTGAAGGGACAGTTCCCGGACACCTACATCCCAACATC	120
Db	27042	AGCTCGCTGGTGGTCTCCGCTTTTAAAGGGGAGCTTTCGTGACACCTACATCCCAACATA	27101
OY	121	GAGGACACCTACCGGGAGGTGATGATCGCGACAAGAGCGGTGCACGTGGCGAGATCACA	180
Db	27102	GAGGACACCTACCGGGAGGTGATGATCGCTGTGCACAGAGCGGTGCACACTGGCAATACAG	27161
OY	181	GACACCAACCGGACAGCCAGTATCCGGGCATGACAGCGCCGTGCATCTCCAAAGGGCAC	240
Db	27162	GACACCAACCGGACAGCCAGTATCCCGGACATGACAGCGGTGCATCTCCAAAGGGCCAC	27221
OY	241	GCCCTCATCCGTTGGTTCCTGCACACAGCAAGTCGCGGAGAGAGCTGGGGGCCATC	300
Db	27222	GCCCTTATCTCGTGTCTCCGTGACACAGCAAGTCGCTGATGAGCTGAGGCCCATC	27281
OY	301	TACAGCTCATCGTCGATCAACAGGGCAGCGTGGAGGACATCCCGGTGATGCTGTGGGC	360





AC103000/c 153386 bp DNA linear HTG 12-JUL-2002  
 LOCUS Rattus norvegicus clone CH230-192M3, \*\*\* SEQUENCING IN PROGRESS  
 DEFINITION \*\*\* 63 unordered pieces.  
 ACCESSION AC103000  
 VERSION AC103000.3 GI:21729632  
 KEYWORDS HTG; HTGS\_PHASE1.  
 SOURCE Norway rat.  
 ORGANISM Rattus norvegicus  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 1 (bases 1 to 153386)  
 Muzny,D.M., Adams,C., Adio-Oduola,B., Alt-osman,F.R., Allen,C., Alshrocks,S.L., Amaralunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbara,J., Benton,J., Blinage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowls,S., Brlewa,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Butrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homi,F., Howard,S., Huber,J., Huliyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisedge,H., Lozdo,R.D., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mahliney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,N., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,N., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogun,M., Okunolu,G., Oetungye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshari,N., Sisson,I., Sodergren,E., Sonake,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Syatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G., and Gibbs,R.

TITLE Direct Submission  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 153386)  
 AUTHORS Worley,K.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (24-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 153386)  
 REFERENCE Direct Submission  
 AUTHORS Worley,K.C.  
 TITLE Submitted (12-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 On Jul 11, 2002 this sequence version replaced gi:17973594.  
 COMMENT ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu

----- Project Information  
 Center project name: GICZ  
 Center clone name: CH230-192M3  
 ----- Summary Statistics  
 Sequencing vector: Plasmid:  
 Chemistry: Dye-terminator Big Dye, 100% of reads  
 Assembly program: Phrap version 0.990329  
 Consensus quality: 83670 bases at least Q40  
 Consensus quality: 86659 bases at least Q30  
 Consensus quality: 88900 bases at least Q20  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 \* NOTE: This is a "working draft" sequence. It currently  
 \* consists of 63 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \*  
 \* 1  
 \* 1030 1029: contig of 1029 bp in length  
 \* 1130 1129: gap of unknown length  
 \* 2363 2363: contig of 1234 bp in length  
 \* 2463 2463: gap of unknown length  
 \* 3463 3463: contig of 1000 bp in length  
 \* 3563 3563: gap of unknown length  
 \* 5122 5122: contig of 1559 bp in length  
 \* 5223 5223: gap of unknown length  
 \* 6502 6502: contig of 1280 bp in length  
 \* 6603 6603: gap of unknown length  
 \* 7941 7941: contig of 1339 bp in length  
 \* 7942 8041: gap of unknown length  
 \* 8042 9124: contig of 1083 bp in length  
 \* 9125 9224: gap of unknown length  
 \* 9225 10411: contig of 1187 bp in length  
 \* 10412 10511: gap of unknown length  
 \* 10512 12121: contig of 1610 bp in length  
 \* 12122 12221: gap of unknown length  
 \* 12222 13926: contig of 1705 bp in length  
 \* 13927 14026: gap of unknown length  
 \* 14027 15525: contig of 1499 bp in length  
 \* 15526 15625: gap of unknown length  
 \* 15626 17072: contig of 1447 bp in length  
 \* 17073 17172: gap of unknown length  
 \* 17173 18485: contig of 1313 bp in length  
 \* 18486 18585: gap of unknown length  
 \* 18586 19821: contig of 1236 bp in length  
 \* 19822 21254: gap of unknown length  
 \* 21255 21354: contig of 1333 bp in length  
 \* 21355 22467: gap of unknown length  
 \* 22468 22567: contig of 1113 bp in length  
 \* 22568 23903: gap of unknown length  
 \* 23904 23904: contig of 1336 bp in length  
 \* 23905 24003: gap of unknown length  
 \* 24004 25789: contig of 1786 bp in length  
 \* 25790 25889: gap of unknown length  
 \* 25890 27325: contig of 1436 bp in length  
 \* 27326 27425: gap of unknown length  
 \* 27426 29320: contig of 1895 bp in length  
 \* 29321 29420: gap of unknown length  
 \* 29421 30862: gap of 1442 bp in length  
 \* 30863 30962: gap of unknown length  
 \* 30963 32929: contig of 1967 bp in length  
 \* 32930 33029: gap of unknown length  
 \* 33030 34612: contig of 1583 bp in length  
 \* 34613 34712: gap of unknown length  
 \* 34713 35663: contig of 1851 bp in length  
 \* 35664 36663: gap of unknown length  
 \* 36664 37847: contig of 1184 bp in length  
 \* 37848 39747: gap of unknown length  
 \* 39748 39747: contig of 1800 bp in length  
 \* 39749 39847: gap of unknown length



*	34864	contig of 1017	bp in length
*	40865	gap of unknown	length
*	40965	contig of 1932	bp in length
*	42897	gap of unknown	length
*	42997	contig of 1214	bp in length
*	44211	gap of unknown	length
*	44311	contig of 1071	bp in length
*	45382	gap of unknown	length
*	45482	contig of 1534	bp in length
*	47016	contig of 1534	bp in length
*	47016	gap of unknown	length
*	47115	contig of 1861	bp in length
*	48996	gap of unknown	length
*	49997	contig of 1413	bp in length
*	50097	contig of 1413	bp in length
*	50510	gap of unknown	length
*	50610	contig of 1338	bp in length
*	51948	gap of unknown	length
*	52047	contig of 2811	bp in length
*	52048	contig of 2811	bp in length
*	54859	gap of unknown	length
*	54959	contig of 1333	bp in length
*	55292	gap of unknown	length
*	55392	contig of 1598	bp in length
*	57990	gap of unknown	length
*	58089	contig of 1624	bp in length
*	58090	contig of 1624	bp in length
*	59714	gap of unknown	length
*	59814	contig of 1691	bp in length
*	61505	gap of unknown	length
*	61605	contig of 1380	bp in length
*	62985	gap of unknown	length
*	63085	contig of 1704	bp in length
*	64788	gap of unknown	length
*	64889	contig of 1694	bp in length
*	65683	gap of unknown	length
*	65683	contig of 2160	bp in length
*	66843	gap of unknown	length
*	68943	contig of 1969	bp in length
*	70932	gap of unknown	length
*	71032	contig of 2302	bp in length
*	73333	gap of unknown	length
*	73433	contig of 2170	bp in length
*	74334	gap of unknown	length
*	75604	contig of 2170	bp in length
*	75704	contig of 2729	bp in length
*	78433	gap of unknown	length
*	78533	contig of 1966	bp in length
*	80529	gap of unknown	length
*	80629	contig of 3054	bp in length
*	83682	gap of unknown	length
*	83783	contig of 4151	bp in length
*	87934	gap of unknown	length
*	88034	contig of 3836	bp in length
*	91870	gap of unknown	length
*	91970	contig of 2907	bp in length
*	94877	gap of unknown	length
*	94977	contig of 5278	bp in length
*	100255	gap of unknown	length
*	100255	contig of 5209	bp in length

Query Match	83.7%;	Score 499.4;	DB 2;	Length 153386;
Best Local Similarity	89.8%;	Pred. No. 1.1e-67;		
Matches 536;	Conservative 0;	Mismatches 61;	Indels 0;	Gaps 0;

QY	1	ATCCGGAAACAGAGTAACGATTAACCGCGGTGGTTCGGGGCGCGCGCGTGGGCAAG	60
Db	83570	ATGCCAGAACAGGAAATGACTACCCGTGGTGTATTTGGGCGAGCGCGGTGGGCAAG	83511
QY	61	AGCTGCTGTGTCCTCGCTTCGTGAAGGACGATTCGGCGACACCTACATCCGACATTC	120
Db	83510	AGCTCCCTGGTBCCTCCGCTTTGTGAAGGGAGCGTTTCGTGACACTTACATCCGACATTA	83451
QY	121	GAGGACACCTACCGCGAGGTGATCGACTTCGCAAGAGCGTGTGCACCGTCGAGATCAGA	180
Db	83450	GAGGACACGTTACCGCGAGGTGATCGACTTCGCAAGAGCGTGTGCACACTCGAGATCAG	83391
QY	181	GACACACCGGCGACCGACCATGTTCCCGGCGCATGCAAGCGCTGTCCATCTCCAAAGGGCCAC	240

Db	83390	GACACCACCGGAGCCACCAGTGTCCCGGCGATGCAACGGCTGTCCATCTTCCAAAGGGCCAC	83331
Qy	241	GCCTTCATCTGGTGGTCTTCGCTCAACCAAGCAAGTCGCTGGAGAGCTGGGGCCCATC	300
Db	83330	GCCTTCATCTGGTGGTCTTCGCGGTGACCAAGCAAGCAAGTCGCTGGAGAGCTGAGGCCCATC	83271
Qy	301	TACAAGCTCATTGTCGATCAATCAAGGGGACGCTGGAGAGATCCCGCTGATGCTGTGGGC	360
Db	83270	TACAAGCTGATGTCGATCAATCAAGGGGACGCTGGAGAGATCCCATCATGCTGTGGGGG	83211
Qy	361	AACAAGTCGATGAACGCAAGCGGAGAGTGGACACGCGGAGGGCAGGCGGTGGCCCA	420
Db	83210	AACAAGTGTGACGAGACGACGCGGAGAGTGGACACCCCGGAGGCGCAGCGCTGGCCACAG	83151
Qy	421	GAGTGAAGTGGGCTTTTCATGAGAGACTTGGCCAAAGTAACTACAACTCAAGAGACTC	480
Db	83150	GAGTGAAGTGGGCTTTCATGAGAGACTTGGCCAAAGTAACTACAACTCAAGAGAGCTG	83091
Qy	481	TTCCAGGAGCTCTACGCTGGAGACGGCGGGAACATGAGCTCAACTCAAGCAGGGCAG	540
Db	83090	TTCCAGGAGCTCTACGCTGGAGACGGCGGGAACATGAGCTCAAGCTCAAGCAGGGCA	83031
Qy	541	CGCTCGGGAACAGAAAGAGACACCGCGTCAAGGGCAAAATGCAACCTCATATGTA	597
Db	83030	CGCTCAACACCAAGAGGGCGGACCGCATCAAGGGCAAGGGCGGCTCATATGTA	82974

RESULT 13					
AB062937					
LOCUS					
DEFINITION	AB062937	4412 bp	mRNA	linear	PRI 13-JUN-2001
	Macaca fascicularis brain cDNA clone:Qf1A-13004, full insert				

VERSION	AB062937.1	GI:14388335
KEYWORDS	oligo capping; f1s (full insert sequence).	
SOURCE	Macaca fascicularis adult male frontal lobe left cDNA to mRNA	

ORGANISM

CCTTCATCCTGATGTTCTCCGTCACCAAGAGAGTCTGTTGAGAGAGCTGGGGCCATC 300  
 CCTTCATCCTGGGTCTCTGCGTACCCAGCAAGAGAGTCTGTGAGAGAGCTGAGCCCATC 83271  
 ACACAGCTCATCTGTGAGATCAAGGGCAGCGCTGGAGAGCATCCCGTATCTCTGTGGC 360  
 ACACAGCTGATTCGTGAGATCAAGGGCAGCGCTGGAGAGCATCCCATATCATCTGTGGGG 83211  
 ACACAGTGCATGTAGAGCAGCGGAGGTGGACACAGCCAGAGCGCAGCGGTGGCCCAA 420  
 ACACAGTGTGACGAGACGACGGGAGGTGCACACACCCTGAGGGCCGACGGCTGGCAAG 83151  
 AGTGAAGATGCGCTTTCATGAGAGACCTCGGCCCAAGATGAACTACACAGTCAGAGAGCTC 480  
 AGTGAAGATGCGCTTTCATGAGAGACCTCGGCCCAAGATGAACTACACAGTCAGAGAGCTC 83091  
 TCCAGAGAGCTGCTGACGCTGGAGACGCGCGGGAACATGACCTTCACATGACGCGCAAG 540  
 TCCAGAGAGCTGCTGACGCTGGAGACGCGCGGGAACATGACCTTCACATGACGCGCAAA 83031  
 GGTCCGCGGGAAGCAGAGAGAGACAGACCGCGCTCAAGGGCAATGACACCTCATGTGA 597  
 GGTCCGCGGGAAGCAGAGAGAGAGACCGCGCTCAAGGGCAATGACACCTCATGTGA 82974  
 AB062937 4412 bp mRNA linear PRI 13-JUN-2001  
 Macaca fascicularis brain cDNA clone:Q1A-13004, full insert  
 sequence.  
 AB062937  
 AB062937.1 GI:14388335  
 oligo capping, fis (full insert sequence).  
 Macaca fascicularis adult male frontal lobe left cDNA to mRNA,  
 clone\_1lb:macaque brain cDNA library Q1A clone:Q1A-13004.  
 Macaca fascicularis  
 Eukaryota: Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 Cercopithecinae; Macaca.  
 1 (sites)  
 Osada, N., Hida, M., Kusuda, J., Tanuma, R., Iseki, K., Hirai, M.,  
 Terao, K., Suzuki, Y., Sugano, S. and Hashimoto, K.  
 Isolation of full-length cDNA clones from macaque brain cDNA  
 libraries  
 unpublished  
 2 (bases 1 to 4412)  
 Hashimoto, K., Osada, N., Hida, M., Kusuda, J. and Sugano, S.  
 Direct Submission  
 Submitted (11-JUN-2001) Katsuyuki Hashimoto, National Institute of  
 Infectious Diseases, Division of Genetic Resources, 23-1, Toyama  
 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan  
 E-mail: khashi@nih.go.jp, URL: http://www.nih.go.jp/yoken/genebank/  
 Tel: 81-3-5285-1111 (ex. 2120), Fax: 81-3-5285-1181  
 Lab host: TOP10  
 Vector: pMR18S-FL3 (Acc. No. AB009864)  
 R. Site1: DraIII (CACTGTGG)  
 R. Site2: DraIII (CACCATGG)  
 Description: 1st strand cDNA was primed with an oligo(dT) primer  
 (ATGCGCCCTTTTCTTTTCTTTT); double-stranded cDNA was synthesized  
 using specific 5' and 3' primers and amplified by PCR. The PCR  
 product was digested with SfiI and size selection was performed to  
 exclude fragments <1.5kb. The SfiI-digested PCR product was cloned  
 into distinct DraIII sites of pMR18S-FL3. XhoI sites just outside  
 the DraIII sites can be used to isolate the cDNA insert. Libraries  
 were constructed by oligo-capping method  
 (Sugano et al., Institute of Medical Science, University of  
 Tokyo).  
 custom primer used for sequencing  
 ( 5' end primer (CTTGTGCTTAAGAGCTCG);  
 3' end primer (CGACTGCAGCTGACGACCA) ).



```

Db 380 TACGACCAATCTGCGACGATCAAGGGAGCGTGGAGAGCATCCCATCATCTGTGTGGG 439
QY 361 AACAGTGGCATGAGACGCG---CGGGAGGTGACACGCGCGAGGGCGAGGGTGGCC 417
Db 440 AACAGTGGCATGAGACGCGCGCGAGGTGACGACGCGAGCGCGAGGGCGTGGCC 499
QY 418 CAAGATGGAAGTGCCTTCATGAGACCTCGGCCAAGATGACTCAACGTCGAAGAG 477
Db 500 CGACATGGAAGTGTGCTTCATGAGACCTCGAGCCAGCTCAACCTTAACGTGAAGAG 559
QY 478 CTCTCCAGAGAGCTGTGACGCTGAGACGCGCGGAACTGAGCCTCAACATCGAGCG 537
Db 560 CTCTCCAGAGAGCTGTGACGCTGAGACGCGCGGAACTGAGCCTCAACATCGAGCG 619
QY 538 AACCGTCCGCGGAAGACGAGACGACGCGGTCAAGGCGCAATGACCTCATGTGA 597
Db 620 AAAAAGACGACGACGAGAAAGAAAGAAAGAAAGCTCAAGAGTGGTATCATGTGA 679

```

RESULT 15  
AB076889

LOCUS AB076889 2505 bp mRNA linear PRI 25-JUN-2002  
DEFINITION Homo sapiens mRNA for DI-Ras2, complete cds.  
ACCESSION AB076889  
VERSION AB076889.1 GI:21624249  
KEYWORDS  
SOURCE Homo sapiens adult brain cDNA to mRNA, clone\_lid:lambda ZAP11 human brain cDNA.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

## REFERENCE

1 Kontani, K., Tada, M., Ogawa, T., Okai, T. and Katada, T.  
DI-Ras: A Distinct Subgroup of Ras-Family GTPases with Unique  
Biochemical Properties  
Unpublished  
2 (bases 1 to 2505)  
Kontani, K., Ogawa, T., Okai, T., Tada, M. and Katada, T.  
Submitted (24-DEC-2001) Toshiaki Katada, University of Tokyo,  
Department of Physiological Chemistry, Graduate School of  
Pharmaceutical Sciences; 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033,  
Japan (E-mail: katada@mol.f.u-tokyo.ac.jp, Tel: 81-3-5841-4750,  
Fax: 81-3-5841-4751)

## JOURNAL

FEATURES  
Source  
1.2505  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="9"  
/map="9q21.33-22.33"  
/tissue\_type="brain"  
/clone\_lid="lambda ZAP11 human brain cDNA"  
/dev\_stage="adult"  
1.2505  
/gene="DI-ras2"  
31..630  
/gene="DI-ras2"  
/note="member of the Ras family  
small GTP-binding protein"  
/codon\_start=1  
/product="DI-Ras2"  
/protein\_id="BAC01116.1"  
/db\_xref="GI:21624250"  
/translation="MPEQSNDRVAVFGAGVGGKSLVLPVKGTFRETYIPVEDTY  
KOVISCDKSICTLTDTGSHQFPMORLSISKHAFILVYSITSRSLEELKPIYE  
QICEIKGVESIPIMLVGNKCDSPSREVOSSFEALARTWCAFMETSAKLHNHYKE  
LEQELINTEKRRVTSLOIDGKKSKOKRREKIKGCVIM"

BASE COUNT 622 a 611 c 630 g 642 t

## ORIGIN

Query Match 63.8%; Score 380.8; DB 9; Length 2505;  
Best Local Similarity 78.3%; Pred. No. 3..3e-49;

```

Matches 470; Conservative 0; Mismatches 127; Indels 3; Gaps 1;
QY 1 ATGCGGACAGAGTACGATTAACCGCTGGTGTTCGGGGCGGGCGGTGGGCAAG 60
Db 31 ATGCTTGAGCAGAGATTACGATTAACCGGTCGCTGTTGGGCTGGCGGTGGGCAAG 90
QY 61 AGCTCGCTGGTGTGCTTCGCTTCGTTGAAAGGACAGTTCGCGACACCTACATCCCAACATC 120
Db 91 AGCTCCCTGGTGTGTTGAGAGTTTGTGAAGGACACATTCGGGAGAGCTACATCCGAGCGTG 150
QY 121 GAGACACCTACCGGCGAGGTGATCAGCTGCACAGAGCTGTGCACGCTGCAGATCACA 180
Db 151 GAAGACACCTACCGGCGAGGTGATCAGCTGCACAGAGCTGTGCACGCTGCAGATCACC 210
QY 181 GACACCGCGGCGAGCGCCACAGTTCGCGCATGAGCGCTGTCATCTCCAAAGGGCAC 240
Db 211 GACACGAGCGGAGCGCCACAGTTCGCGCATGAGCGCGCTGTCATCTCCAAAGGGCAC 270
QY 241 GCCTCATCTGCTGGTGTCTCCGTCACGACGACGACGCTGAGAGCTGGGGCCCATC 300
Db 271 GCCTCATCTGCTGGTGTCTCCGTCACGACGACGACGCTGAGAGCTGGGGCCCATC 330
QY 301 TACAGCTCATCTGTCAGATCAAGGCGAGCGTGGAGGACATCCCGTATCTCTGTGGGC 360
Db 331 TACGACCAATCTGCGAGATCAAGGCGAGCGTGGAGGACATCCCGTATCTCTGTGGGC 390
QY 361 AACAGTGGCATGAGACGCG---CGGGAGGTGACACGCGCGAGGGCGAGGGTGGCC 417
Db 391 AACAGTGGCATGAGACGCGCGCGAGCGGTGACGACGCGAGGGCGAGGGTGGCC 450
QY 418 CAAGATGGAAGTGCCTTCATGAGACCTCGGCCAAGATGACTCAACGTCGAAGAG 477
Db 451 CGCAGATGGAAGTGTGCTTCATGAGACCTCGAGCCAGCTCAACCTTAACGTGAAGAG 510
QY 478 CTCTCCAGAGAGCTGTGACGCTGAGACGCGCGGAACTGAGCCTCAACATCGAGCGC 537
Db 511 CTCTCCAGAGAGCTGTGACGCTGAGACGCGCGGAACTGAGCCTCAACATCGAGCGC 570
QY 538 AACCGTCCGCGGAAGACGAGACGACGCGGTCAAGGCGCAATGACCTCATGTGA 597
Db 571 AAAAAGACGACGACGAGAAAGAAAGAAAGAAAGCTCAAGAGTGGTATCATGTGA 630

```

Search completed: March 23, 2003, 16:09:55  
Job time: 2385.86 secs



GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: March 23, 2003, 13:56:56 ; Search time 72.4223 Seconds  
(without alignments)  
8438.824 Million cell updates/sec

Title: US-09-873-546-6  
Perfect score: 21  
Sequence: 1 gcgcggcgaagaactcgcctg 21

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapept 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

## Database :

1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rod:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vtl:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19.4	92.4	597	9 AY056037	AY056037 Homo sapi
2	19.4	92.4	597	9 AY059641	AY059641 Homo sapi
3	19.4	92.4	597	10 AY059640	AY059640 Mus muscu
4	19.4	92.4	774	9 AB076888	AB076888 Homo sapi
5	19.4	92.4	1249	6 AX430295	AX430295 Sequence
6	19.4	92.4	1574	10 AB008423	AB008423 Rattus no
7	19.4	92.4	1624	10 RACGYPB2	M16655 Rat cytochr
8	19.4	92.4	1698	10 RACGYPB2	M22330 Rat cytochr
9	19.4	92.4	2763	1 CCU56652	U56652 Caulobacter
10	19.4	92.4	2827	9 AK096600	AK096600 Homo sapi
11	19.4	92.4	3061	6 AX430418	AX430418 Sequence
12	19.4	92.4	3391	9 BC030660	BC030660 Homo sapi
13	19.4	92.4	6710	1 CAJ10321	CAJ10321 Caulobact
14	19.4	92.4	7701	10 RS1ID26	X52027 Rattus norv
15	19.4	92.4	10960	1 AE005869	AE005869 Caulobact
16	19.4	92.4	126901	2 AC091518	AC091518 Mus muscu
17	19.4	92.4	176340	2 AC120750	AC120750 Rattus no
18	19.4	92.4	177540	9 AC006538	AC006538 Homo sapi
19	19.4	92.4	180544	2 AC107527	AC107527 Rattus no
20	19.4	92.4	201320	2 AC073816	AC073816 Mus muscu
21	18.4	87.6	633	1 AF180424	AF180424 Helicobac
22	18.4	87.6	757	1 AF315501	AF315501 Helicobac
23	17.8	84.8	278	1 AF320123	AF320123 Mycobacte
24	17.8	84.8	1034	3 AF255672	AF255672 Drosophil
25	17.8	84.8	1218	9 HSP450	X16865 Human mRNA
26	17.8	84.8	1449	6 AX411044	AX411044 Sequence
27	17.8	84.8	1449	9 HSDBI1A	X07618 Human mRNA
28	17.8	84.8	1450	6 AX192411	AX192411 Sequence
29	17.8	84.8	1494	6 E10647	E10647 Human cDNA
30	17.8	84.8	1494	6 E10867	E10867 cDNA encodi
31	17.8	84.8	1494	6 E10868	E10868 cDNA encodi
32	17.8	84.8	1494	6 E10869	E10869 cDNA encodi
33	17.8	84.8	1494	6 E10870	E10870 cDNA encodi
34	17.8	84.8	1494	6 MFU38218	U38218 Macaca fasc
35	17.8	84.8	1545	6 E15820	E15820 cDNA encodi
36	17.8	84.8	1563	9 HSP450IID	X16867 Human mRNA
37	17.8	84.8	1566	6 AR084365	AR084365 Sequence
38	17.8	84.8	1566	6 AR084374	AR084374 Sequence
39	17.8	84.8	1567	9 HSDBI1	X08006 Homo sapien
40	17.8	84.8	1567	9 HUMCYPDB1	M20403 Human cyloc
41	17.8	84.8	1568	6 A20907	A20907 debriIsoquin
42	17.8	84.8	1568	6 AR084372	AR084372 Sequence
43	17.8	84.8	1569	6 AR084375	AR084375 Sequence
44	17.8	84.8	1571	6 AR084373	AR084373 Sequence
45	17.8	84.8	1593	6 AX195173	AX195173 Sequence

## ALIGNMENTS

RESULT 1  
LOCUS AY056037 597 bp mRNA linear PRI 24-JUL-2002  
DEFINITION Homo sapiens Rtg protein mRNA, complete cds.  
ACCESSION AY056037  
VERSION AY056037.1 GI:16555333  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 597)  
AUTHORS Ellis,C.A., Vos,M.D., Howell,H., Vallecora,T., Puls,D.W. and Clark,G.J.  
TITLE Rtg is a novel Ras-related protein and potential neural tumor

JOURNAL  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source

Suppressor  
Proc. Natl. Acad. Sci. U.S.A. 99 (15), 9876-9881 (2002)  
12107278  
2 (bases 1 to 597)  
Castro, A.F. and Quilliam, L.A.  
Identification of Rlg, a novel Ras family member sharing homology  
with NOEY2 and Rap  
Unpublished  
3 (bases 1 to 597)  
Castro, A.F. and Quilliam, L.A.  
Direct Submission  
Submitted (13-SEP-2001) Biochemistry and Molecular Biology, Indiana  
University School of Medicine, 635 Barnhill Drive, MS-4053,  
Indianapolis, IN 46202, USA  
Location/Qualifiers  
1. .597  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="19"  
/map="19p13.3"  
/tissue\_type="brain; anaplastic oligodendroglioma"  
/note="derived from dbEST A1497811"  
1. .597  
/note="Ras family GTPase"  
/codon\_start=1  
/product="Rlg protein"  
/protein\_id="AAL23715.1"  
/db\_xref="GI:16535334"  
/translation="MPEQSNDRVAVFVGAGVGSLSVLRFKVGFPRDTPYIPEDTY  
RQVISCDSKSVCTLOITDTTGSHQFPAMORLSISKGHAFILFVSYSKOSLELGPYIK  
LIVQIKGSVEDIPVLVGNKCDREVDTRDAQVADQWKAFCAFMETSANKMYNVEL  
FOELLTLETRRMSLNIDGKRSKGRKRDYVKGCTLM"

BASE COUNT 134 a 178 c 195 g 90 t

ORIGIN

Query Match 92.4%; Score 19.4; DB 9; Length 597;  
Best Local Similarity 95.2%; Pred. No. 53;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCGTGGCAAGACTGCTGG 21  
|||||  
Db 50 GCGTGGCAAGACTGCTGG 70

RESULT 2  
AY059641  
LOCUS 597 bp mRNA linear PRI 29-OCT-2001  
DEFINITION Homo sapiens small GTP-binding tumor suppressor 1 mRNA, complete  
cds.  
ACCESSION AY059641  
VERSION AY059641.1 GI:16508175  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 597)  
Gong, L. and Wu, K.  
Molecular cloning of GBRSL, a novel gene encoding a small  
GTP-binding tumor suppressor  
Unpublished  
2 (bases 1 to 597)  
Gong, L.  
Direct Submission  
Submitted (12-OCT-2001) Dept. of Cardiology, U.T. MD. Anderson  
Cancer Center, 1515 Holcombe Blvd-449, Houston, TX 77030, USA  
Location/Qualifiers  
1. .597  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="19"  
1. .597  
/note="GBRSL"

CDS

FEATURES  
source

/codon\_start=1  
/product="small GTP-binding tumor suppressor 1"  
/protein\_id="AAL17968.1"  
/db\_xref="GI:16508176"  
/translation="MPEQSNDRVAVFVGAGVGSLSVLRFKVGFPRDTPYIPEDTY  
RQVISCDSKSVCTLOITDTTGSHQFPAMORLSISKGHAFILFVSYSKOSLELGPYIK  
LIVQIKGSVEDIPVLVGNKCDREVDTRDAQVADQWKAFCAFMETSANKMYNVEL  
FOELLTLETRRMSLNIDGKRSKGRKRDYVKGCTLM"

BASE COUNT 134 a 178 c 195 g 90 t

ORIGIN

Query Match 92.4%; Score 19.4; DB 9; Length 597;  
Best Local Similarity 95.2%; Pred. No. 53;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCGTGGCAAGACTGCTGG 21  
|||||  
Db 50 GCGTGGCAAGACTGCTGG 70

RESULT 3  
AY059640  
LOCUS 597 bp mRNA linear ROD 29-OCT-2001  
DEFINITION Mus musculus small GTP-binding tumor suppressor 1 (Gbrsl) mRNA,  
complete cds.  
ACCESSION AY059640  
VERSION AY059640.1 GI:16508173  
KEYWORDS  
SOURCE Mus musculus.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 597)  
Gong, L. and Wu, K.  
Molecular cloning of GBRSL, a novel gene encoding a small  
GTP-binding tumor suppressor  
Unpublished  
2 (bases 1 to 597)  
Gong, L.  
Direct Submission  
Submitted (12-OCT-2001) Dept. of Cardiology, U.T. MD. Anderson  
Cancer Center, 1515 Holcombe Blvd-449, Houston, TX 77030, USA  
Location/Qualifiers  
1. .597  
/organism="Mus musculus"  
/strain="129/SvJ"  
/db\_xref="taxon:10090"  
1. .597  
/gene="Gbrsl"  
/gene="Gbrsl"  
/gene="Gbrsl"  
/note="GBRSL"  
/codon\_start=1  
/product="small GTP-binding tumor suppressor 1"  
/protein\_id="AAL17967.1"  
/db\_xref="GI:16508174"  
/translation="MPEQSNDRVAVFVGAGVGSLSVLRFKVGFPRDTPYIPEDTY  
RQVISCDSKSVCTLOITDTTGSHQFPAMORLSISKGHAFILFVSYSKOSLELGPYIK  
LIVQIKGSVEDIPVLVGNKCDREVDTRDAQVADQWKAFCAFMETSANKMYNVEL  
FOELLTLETRRMSLNIDGKRSKGRKRDYVKGCTLM"

BASE COUNT 138 a 166 c 194 g 99 t

ORIGIN

Query Match 92.4%; Score 19.4; DB 10; Length 597;  
Best Local Similarity 95.2%; Pred. No. 53;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCGTGGCAAGACTGCTGG 21  
|||||  
Db 50 GCGTGGCAAGACTGCTGG 70

RESULT 4

AB076888  
LOCUS AB076888 774 bp mRNA linear PRI 25-JUN-2002  
DEFINITION Homo sapiens mRNA for Di-Ras1, complete cds.  
ACCESSION AB076888  
VERSION AB076888.1 GI:21624247  
KEYWORDS  
SOURCE Homo sapiens adult brain cDNA to mRNA, clone\_l1b:lambda ZAPII human brain cDNA.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Kontani,K., Tada,M., Ogawa,T., Okai,T. and Katada,T.  
TITLE Di-Ras: A Distinct Subgroup of Ras-family GTPases with Unique Biochemical Properties  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 774)  
AUTHORS Kontani,K., Ogawa,T., Okai,T., Tada,M. and Katada,T.  
TITLE Direct Submission  
JOURNAL Submitted (24-DEC-2001) Toshiaki Katada, University of Tokyo, Department of Physiological Chemistry, Graduate School of Pharmaceutical Sciences, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan (E-mail:katadaemol.t.u-tokyo.ac.jp, Tel:81-3-5841-4750, Fax:81-3-5841-4751)  
FEATURES  
source Location/Qualifiers  
1..774  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="19"  
/map="19p13.3"  
/tissue\_type="brain"  
/clone\_lib="lambda ZAPII human brain cDNA"  
/dev\_stage="adult"  
1..774  
/gene="Di-ras1"  
100..696  
/gene="Di-ras1"  
/note="member of the Ras family small GTP-binding protein"  
/codon\_start=1  
/product="Di-Ras1"  
/protein\_id="BAC01115.1"  
/db\_xref="GI:21624248"  
/translation="MPEQSNDRVVFAGVGKSLVLRFGKTFRDYIPTIEDYROYISCKSVCTLOITDPTGSHOPAPARQISLGHFLIVFSTYSQSLLELPITKLYIQIKGVSVDIPYMLGVKNCDFQREVDREARAVAGQEMKCAFMETSAKNNYVKELFOELTETRNMSLNIDGKRSKOKRTDRYKCKTLT"  
BASE COUNT 154 a 268 c 239 g 113 t  
ORIGIN  
Query Match 92.4%; Score 19.4; DB 9; Length 774;  
Best Local Similarity 95.2%; Pred. No. 53;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GCGTGGCAGACACTCGCTGG 21  
|||||  
Db 149 GCGTGGCAGACAGCTCGCTGG 169  
RESULT 5  
AX430295 1249 bp DNA linear PAT 28-JUN-2002  
LOCUS AX430295  
DEFINITION Sequence 77 from Patent W00240715.  
ACCESSION AX430295  
VERSION AX430295.1 GI:21655659  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Chalup,M.S., Altus,C.M., Lincoln,S.E., Dufour,G.E. and Jackson,S.  
TITLE Molecules for disease detection and

JOURNAL Patent: WO 0240715-A 77 23-MAY-2002;  
INCYTE GENOMICS INC (US)  
FEATURES Location/Qualifiers  
source 1..1249  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/note="IncYTE ID No: LG:389870.1:2000SEP08"  
BASE COUNT 230 a 436 c 355 g 228 t  
ORIGIN  
Query Match 92.4%; Score 19.4; DB 6; Length 1249;  
Best Local Similarity 95.2%; Pred. No. 54;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GCGTGGCAGACACTCGCTGG 21  
|||||  
Db 100 GCGTGGCAGACACTCGCTGG 120  
RESULT 6  
AB008423 1574 bp mRNA linear ROD 29-OCT-1997  
LOCUS AB008423  
DEFINITION Rattus norvegicus mRNA for CYP2D2, complete cds.  
ACCESSION AB008423  
VERSION AB008423.1 GI:2575858  
KEYWORDS CYP2D2.  
SOURCE Rattus norvegicus (strain:Sprague-Dawley) 7 weeks age male liver  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
REFERENCE  
AUTHORS Wan,J., Imaoka,S., Chow,T., Hiroi,T., Yabusaki,Y. and Funae,Y.  
TITLE Expression of four rat CYP2D isoforms in Saccharomyces cerevisiae and their catalytic specificity  
JOURNAL Arch. Biochem. Biophys. (1997) In press  
AUTHORS 2 (bases 1 to 1574)  
TITLE Imaoka,S.  
JOURNAL Direct Submission.  
TITLE Submitted (22-OCT-1997) Susumu Imaoka, Osaka City University Medical School, Laboratory of Chemistry; 1-4-54 Asahimachi,, Abeno-ku, Osaka 545, Japan (E-mail:Imaoka@msc.med.osaka-cu.ac.jp, Tel:81-6-645-2081, Fax:81-6-646-3922)  
FEATURES  
source Location/Qualifiers  
1..1574  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/sex="male"  
/tissue\_type="liver"  
/dev\_stage="7 weeks age"  
/tissue\_lib="cDNA"  
1..1574  
/gene="Cyp2d2"  
1..1503  
/gene="Cyp2d2"  
/note="CYP2D cytochrome P450 isoform"  
/codon\_start=1  
/product="CYP2D2"  
/protein\_id="BAA23123.1"  
/db\_xref="GI:2575859"  
/translation="MGILIGDDIANAVYFTAILIIVLYVHRHKFWTAHYPRGPVPLPGLNDLVQDFENMPYSLYKLRSGRGDFSLDIANKPVVYINGLAVRRLVYTGEDTA DRPLIYNIHLYGYNKSKGVVLAPGPEWRQRRSVSTLDFGVGKSLQWVTEEA GHLDQTFARKAEHPNPISLISKAVSNVIAVLRREYEDPFNNRLKTLKSFEE DTGMAEVLNAILPILQIPGLPGKGVKPLNFEIALVDMLEHKKSMPPADPPDMVD AFLAMOKAKNPSPSSNDENLRIVVIDLFAAGVYTSSTLSMAILLIILHPPQRYV HEETIDVYGVRRPEMDAOPFTNNAYIHFVORADIVPNTIHMTMSRDIKQGGFLI PKGTTLLPNSSVYKDETVMKPLRFPHRHHLDQGNFVKHEAPHSAGRACLGEPLARMELEFETFCLLQRFSESVLAKRPSTHGVYALPVPYDOLCAVAV"  
BASE COUNT 343 a 456 c 416 g 359 t  
ORIGIN

Query Match 92.4%; Score 19.4; DB 10; Length 1574;  
Best Local Similarity 95.2%; Pred. No. 54;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GCGTGGCAAGAACTCGCTGG 21  
|||||  
Db 437 GCGTGGCAAGAAATCGCTGG 457

RESULT 7  
RATCYPDB2 1624 bp mRNA linear ROD 27-Apr-1993  
LOCUS Rat cytochrome P450-db2 mRNA, complete cds.  
DEFINITION M16655.1 GI:203835  
ACCESSION cytochrome P450; debrisoquine 4-hydroxylase.  
VERSION Rat (Sprague-Dawley, 8 week old) liver microsomal, cDNA to mRNA,  
KEYWORDS clone pP450db2.  
SOURCE Rattus norvegicus  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 1624)  
Gonzalez, F.J., Matsunaga, T., Nagata, K., Meyer, U.A., Nebert, D.W.,  
Pastewka, J., Kozak, C.A., Gillette, J., Gelboin, H.V. and  
Hardick, J.P.  
Debrisoquine 4-hydroxylase: characterization of a new P450 gene  
subfamily, regulation, chromosomal mapping, and molecular analysis  
of the DA rat polymorphism  
DNA 6 (2), 149-161 (1987)  
JOURNAL MEDLINE 87217961  
PUBMED 3582092

FEATURES  
source  
location/Qualifiers  
1..1624  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
<1..1624  
/product="CYP450 mRNA"  
51..1553  
/note="debrisoquine 4-hydroxylase."  
/codon\_start=1  
/protein\_id="AA41055.1"  
/db\_xref="GI:203835"  
/translation="MGILIGDDLMNAVITAFLLVLDLVHNRKENTAHYPPGPPPLP  
GGLNLDVDFENMPYSLYKLRSGYGVESLQIAKRVVYVINGKARRELLVYGEDTA  
DRPLPIYNHIGYGNKSGVLAAPGPERDRRSVTLRPGVAKSLQEDWTEEA  
GHLCDFEAKAEHPNPSTILLSKAVSNVYASLYARFEYEDPFENRMKLTKESEGE  
DTGFMAEVNLAIPILQIGLPGKVPKINSFTALVDKMLIEHKKSWDPAOPPRDRTD  
AFLEAKOKAKGNPESSEFNDENLRLVVIDLFMGWVTTSTLSMALLMLHLDVQRRV  
HEIDEVIGQVRRPEMADQARMPLTNATVHEVORADIVPTNIPHTSKDIFQGFLLI  
PKGTTLIPNLSSVLDDEYWEKPLRPHPHFLDQGNFVKAHEAFMFSAGRACIGEP  
LARMELFLFETCLORFSVLAGRPRSTHGVALPVPPOYOLCAVAR"

BASE COUNT 347 a 466 c 437 g 374 t  
ORIGIN Unreported.

Query Match 92.4%; Score 19.4; DB 10; Length 1624;  
Best Local Similarity 95.2%; Pred. No. 54;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GCGTGGCAAGAACTCGCTGG 21  
|||||  
Db 487 GCGTGGCAAGAAATCGCTGG 507

RESULT 8  
RATCYPC 1698 bp mRNA linear ROD 27-Apr-1993  
LOCUS Rat cytochrome P450 isozyme CME2 mRNA, complete cds.  
DEFINITION M22330.1 GI:203823  
ACCESSION cytochrome P450.  
VERSION Rat (Sprague-Dawley, male) liver, cDNA to mRNA, clone

ORGANISM lambda-circME[22.1].  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 1698)  
Ishida, N., Tawaragi, Y., Inuzuka, C., Sugita, O., Kubota, I.,  
Nakazato, H., Noguchi, T. and Sassa, S.  
Four species of cDNAs for cytochrome P450 isozymes immunorelated to  
P450C-M/E encode for members of P450IId subfamily, increasing the  
number of members within the subfamily  
Biochem. Biophys. Res. Commun. 156 (2), 681-688 (1988)  
JOURNAL MEDLINE 89050091  
PUBMED 3190674

COMMENT Draft entry and computer-readable sequence for [1] kindly provided  
by N. Ishida, 20-JAN-1989.  
location/Qualifiers  
1..1698  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
<1..1698  
/product="CYP mRNA"  
125..1627  
/note="cytochrome P450"  
/codon\_start=1  
/protein\_id="AA41049.1"  
/db\_xref="GI:203824"  
/translation="MGILIGDDLMNAVITAFLLVLDLVHNRKENTAHYPPGPPPLP  
GGLNLDVDFENMPYSLYKLRSGYGVESLQIAKRVVYVINGKARRELLVYGEDTA  
DRPLPIYNHIGYGNKSGVLAAPGPERDRRSVTLRPGVAKSLQEDWTEEA  
GHLCDFEAKAEHPNPSTILLSKAVSNVYASLYARFEYEDPFENRMKLTKESEGE  
DTGFMAEVNLAIPILQIGLPGKVPKINSFTALVDKMLIEHKKSWDPAOPPRDRTD  
AFLEAKOKAKGNPESSEFNDENLRLVVIDLFMGWVTTSTLSMALLMLHLDVQRRV  
HEIDEVIGQVRRPEMADQARMPLTNATVHEVORADIVPTNIPHTSKDIFQGFLLI  
PKGTTLIPNLSSVLDDEYWEKPLRPHPHFLDQGNFVKAHEAFMFSAGRACIGEP  
LARMELFLFETCLORFSVLAGRPRSTHGVALPVPPOYOLCAVAR"

BASE COUNT 373 a 483 c 450 g 392 t  
ORIGIN 311 bp upstream of HindIII site.

Query Match 92.4%; Score 19.4; DB 10; Length 1698;  
Best Local Similarity 95.2%; Pred. No. 54;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GCGTGGCAAGAACTCGCTGG 21  
|||||  
Db 561 GCGTGGCAAGAAATCGCTGG 581

RESULT 9  
CCU56652 2763 bp DNA linear BCF 14-NOV-1996  
LOCUS Caulobacter crescentus lon protease (lon) gene, complete cds.  
DEFINITION CCU56652  
ACCESSION U56652  
VERSION U56652.1 GI:1336062

KEYWORDS  
SOURCE Caulobacter crescentus.  
ORGANISM Caulobacter crescentus  
Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;  
Caulobacter.  
1 (bases 1 to 2763)  
Wright, R., Stephens, C., Zweiger, G., Shapiro, L. and Alley, M.R.  
Caulobacter lon protease has a critical role in cell-cycle control  
of DNA methylation  
Genes Dev. 10 (12), 1532-1542 (1996)  
JOURNAL MEDLINE 96258253  
PUBMED 8666236

REFERENCE 2 (bases 1 to 2763)  
Wright, R., Stephens, C., Zweiger, G., Shapiro, L. and Alley, M.R.  
Direct Submission  
Submitted (24-Apr-1996) Developmental Biology, Stanford University,  
Beckman Center B300, Stanford Medical Center, Stanford, CA  
94305-5427, USA  
location/Qualifiers



```

source
1. .2763
/organism="Caulobacter crescentus"
/strain="NA1000"
/db_xref="taxon:155892"
340. .2740
/gene="lon"
340. .2739
/gene="lon"
/codon_start=1
/transl_table=11
/product="lon protease"
/protein_id="AAB18765.1"
/db_xref="GI:166739"
/translacion="MSELRLPLVPLRLDIYVEPHWVPLFVGRDYSVLEVMRGDK
OILLVTKNSADDDPAPGDIFFGVIAVLIOLLPCTVTVIVGKRAAVVSTGTO
ESTYEAQIGEVSEDDGAPPEAGLSRAVEDPENVYKINKVPEPALSTIPIAPGK
LADSIKRAHLSVIGDKONLEIFDVVKLEVFVLMSEISLVLEVERKIRSVKROME
KTOREYVLNEQMKAIQRELSDPDARDELIDLERIKKTKLSKEARTKASEELKLN
MSPMSAEVTVRNKLDWLLSIPMGKAKTKIDLVSEGIILADHYGLEKVERILEYL
AVOARTNSLKGPILCLVGPVGKTSIAKADGREFVMSIGGVGLVEDEIRGHR
TYISMGKVVQSMKKAKTNAFVLDEIDMGSDRGDPASALLEVDPSQNSFCD
HYLEVDYELSQVMEVTTANSLNMPQLIDRMETIRIPCTYDEKLEIAKRHILEPLAK
DHGLKPAEFIVPDKAIRDLRYVTRAGVRSLEBELGALKATVADLAREKVASITID
DERLAKYAGVKKRYGETVDQVGIYVGLMTEFGDILITEAKMKGKGMQITGN
LKDYAKESIAANSYVRSALQFGIKPVEKPTVYHIHPDGATPKDPSAGIAMALA
MSVLTGIPIRKDIAMGETILRGRTAIGGLKEKLLAALRSVKTVLIPQENEDLA
DVPQTVKDGLEIIFVSTYDEVKLHNLGRLTPVEMNEAEPTTSAKKDDSDAMLT
H"

BASE COUNT      564 a      817 c      923 g      459 t
ORIGIN
Query Match      92.4%; Score 19.4; DB 1; Length 2763;
Best Local Similarity 95.2%; Pred. No. 55;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCGTGGGCAAGACTCGCTGG 21
|||||
Db 1403 GCGTGGCAAGACTCGCTGG 1423

RESULT 10
AK096600      2827 bp mRNA linear PRI 15-JUL-2002
LOCUS      Homo sapiens cDNA FLJ39281 fls, clone OCBBF2011067, highly similar
to Homo sapiens cingulin mRNA.
ACCESSION      AK096600.1 GI:21756131
VERSION      AK096600.1
KEYWORDS      oligo capping; fits (full insert sequence).
SOURCE      Homo sapiens fetal brain cDNA to mRNA, clone_lib:OCBBF2
clone:OCBBF2011067.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 Nishi,T., Ota,T., Nakagawa,S., Senoh,A., Mizuguchi,H., Inagaki,H.,
Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,
Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,
Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,
Kikuchi,H., Kanda,K., Wagatsuna,M., Murakawa,K., Kanehori,K.,
Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y.,
Sugano,S., Nagahari,K., Masubo,Y., Nagai,K. and Isogai,T.,
NEBO human cDNA sequencing project
Unpublished
2 (bases 1 to 2827)
Isogai,T. and Yamamoto,J.
Direct Submission
Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team); 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomicehri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEBO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan

```

```

FEATURES
source
1. .2827
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="OCBBF2011067"
/tissue_type="brain"
/clone_lib="OCBBF2"
/dev_stage="fetal"
/note="Cloning vector: pME18SFJ3"

BASE COUNT      688 a      750 c      983 g      406 t
ORIGIN
Query Match      92.4%; Score 19.4; DB 9; Length 2827;
Best Local Similarity 95.2%; Pred. No. 55;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCGTGGGCAAGACTCGCTGG 21
|||||
Db 179 GCGTGGGCAAGACTCGCTGG 199

RESULT 11
AX430418      3061 bp DNA linear PAT 28-JUN-2002
LOCUS      AX430418
DEFINITION      Sequence 200 from Patent WO240715.
ACCESSION      AX430418
VERSION      AX430418.1 GI:21655782
KEYWORDS      human.
SOURCE      Homo sapiens
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 Chalup,M.S., Altus,C.M., Lincoln,S.E., Dufour,G.E. and Jackson,S.
AUTHORS      Molecules for disease detection and treatment
TITLE      Patent: WO 0240715-A 200 23-MAY-2002;
JOURNAL      INCYTE GENOMICS INC (US)
FEATURES
source
1. .3061
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="Incyle ID No: LI:332161.1:2000SEP08"

BASE COUNT      606 a      840 c      927 g      688 t
ORIGIN
Query Match      92.4%; Score 19.4; DB 6; Length 3061;
Best Local Similarity 95.2%; Pred. No. 55;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCGTGGGCAAGACTCGCTGG 21
|||||
Db 100 GCGTGGGCAAGACTCGCTGG 120

RESULT 12
BC030660      3391 bp mRNA linear PRI 21-MAY-2002
LOCUS      BC030660
DEFINITION      Homo sapiens, similar to R1g protein, clone MGC:33391
IMAGE:4814337, mRNA, complete cds.
ACCESSION      BC030660
VERSION      BC030660.1 GI:21040534
KEYWORDS      MGC.
SOURCE      Homo sapiens
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 3391)
AUTHORS      Strausberg,R.
TITLE      Direct Submission

```

## JOURNAL

Submitted (20-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

## REMARK COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
 CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN)  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 CDNA Sequencing by: Institute for Systems Biology  
<http://www.systemsbio.org>  
 contact: [amadesystemsbiology.org](mailto:amadesystemsbiology.org)  
 Anup Madan, Jessica Fahney, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
 Series: IRK Plate: 46 Row: e Column: 8  
 This clone was selected for full length sequencing because it passed the following selection criteria: Genomescan gene prediction.

## FEATURES

## source

Location/Qualifiers  
 1..3391  
 /organism="Homo sapiens"  
 /db\_xref="LocusID:148252"  
 /db\_xref="taxon:9606"  
 /clone="MGC:33391 IMAGE:4814337"  
 /tissue\_type="Brain, hippocampus"  
 /clone\_1lb="NIH\_MGC\_95"  
 /lab\_host="DH10B"  
 /note="Vector: pBluescript"  
 138..734  
 /codon\_start=1  
 /product="similar to Rlg protein"  
 /protein\_id="AAH30660.1"  
 /db\_xref="GI:21040535"  
 /translation="MEQSNDRVYVFGAGVGKSLVLPVKGFRTDYPTIEDTY ROVISCDKSVCTLIQITDTGSHQFPAMORLISKGAFILVSVTSKQSLBGLPIYK LIVOIKGSVEDIPVMLVGNKCDTQREVDTREAOAVAOEMKCAFMTSAKMVYVKEL FQELLTLERRMSLINIDGKRSKGRKTRDRAVKGKCTLM"

## CDS

BASE COUNT 650 a 996 c 1016 g 729 t  
 ORIGIN

Query Match 92.4%; Score 19.4; DB 9; Length 3391;  
 Best Local Similarity 95.2%; Pred. No. 55;

Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 GCCTGGCGCAAGACTGCTGG 21

|||||

Db 187 GCCTGGCGCAAGACTGCTGG 207

## RESULT 13

## CAJ10321

LOCUS CAJ10321 6710 bp DNA linear BCT 10-MAR-2001  
 DEFINITION Caulobacter crescentus partial tlg gene and clpP, cica, clpX, lon genes.

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

1 (bases 1 to 6710)  
 Fuchs,T., Wiget,P., Osteras,M. and Jenal,U.  
 Precise amounts of a novel member of a phosphotransferase superfamily are essential for growth and normal morphology in Caulobacter crescentus  
 Mol. Microbiol. 39 (3), 679-692 (2001)

PUBMED 11169108  
 REFERENCE 2 (bases 1 to 6710)  
 AUTHORS Jenal,U.

## JOURNAL

## TITLE

## JOURNAL

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE

## TITLE



Intron 4205..4304 /gene="P450 IID2"  
/number=8  
4305..>4483 /gene="P450 IID2"  
/number=9  
/function="codes for a region of P450 protein that is  
associated with the noncovalently bound heme iron at the  
enzymes active site."  
polya\_signal 4533..4538  
/gene="P450 IID2"  
4534  
polya\_site /gene="P450 IID2"  
BASE COUNT 2039 a 1783 c 2019 g 1859 t 1 others  
ORIGIN

Query Match 92.4%; Score 19.4; DB 10; Length 7701;  
Best Local Similarity 95.2%; Pred. No. 57;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 GCCTGGCAAGACTGCTGG 21  
|||||  
Db 1874 GCCTGGCAAGAAATGCTGG 1894

RESULT 15  
LOCUS AE005869/c  
DEFINITION Caulobacter crescentus CB15 section 195 of 359 of the complete  
genome.  
ACCESSION AE005869  
VERSION AE005869.1 GI:13423417  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

1 (bases 1 to 10960)  
Nierman, W.C., Feldblyum, T.V., Laub, M.T., Paulsen, I.T., Nelson, K.E.,  
Eisen, J., Heidelberg, J.F., Alley, M.R., Ohta, N., Maddock, J.R.,  
Potocka, I., Nelson, W.C., Newton, A., Stephens, C., Phadke, N.D.,  
Ely, B., Debroy, R.T., Dodson, R.J., Durkin, A.S., Gwinn, M.L.,  
Halt, D.H., Kolonay, J.F., Smit, J., Craven, M.B., Khouri, H.,  
Shetty, J., Berry, K., Uterback, T., Tran, K., Wolf, A., Vamathevan, J.,  
Ermoлаeva, M., White, O., Salzberg, S.L., Venter, J.C., Shapiro, L. and  
Fraser, C.M.  
Complete genome sequence of *Caulobacter crescentus*  
Proc. Natl. Acad. Sci. U.S.A. 98 (7), 4136-4141 (2001)  
21173698  
11259647

2 (bases 1 to 10960)  
Nierman, W.C., Feldblyum, T.V., Paulsen, I.T., Nelson, K.E., Eisen, J.,  
Heidelberg, J.F., Alley, M.R., Ohta, N., Maddock, J.R., Potocka, I.,  
Nelson, W.C., Newton, A., Stephens, C., Phadke, N.D., Ely, B.,  
Laub, M.T., Debroy, R.T., Dodson, R.J., Durkin, A.S., Gwinn, M.L.,  
Halt, D.H., Kolonay, J.F., Smit, J., Craven, M., Khouri, H., Shetty, J.,  
Berry, K., Uterback, T., Tran, K., Wolf, A., Vamathevan, J.,  
Ermoлаeva, M., White, O., Salzberg, S.L., Shapiro, L., Venter, J.C. and  
Fraser, C.M.  
Direct Submission  
Submitted (31-JAN-2001) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA  
Location/Qualifiers  
1. 10960  
/organism="Caulobacter crescentus CB15"  
/strain="CB15"  
/db\_xref="taxon:190650"  
131..207  
/gene="trna-asp-1"  
131..207  
/gene="trna-asp-1"  
/product="trna-asp"  
272..348  
gene

trna /gene="trna-asp-2"  
272..348  
/gene="trna-asp-2"  
/product="trna-asp"  
407..1420  
/gene="CC1957"  
407..1420  
/gene="CC1957"  
/note="identified by glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAK23932.1"  
/db\_xref="GI:13423418"  
/translation="MRRLADMTSAALPPISVLDLAPVPGSSVGGALRNSIDLACAAE  
TLGFHRYVLAHHNMGPGLASAAATVAVVIGIATTTIRVSGSGVMLPNHAPLVAEAF  
GTLNMLPGRIDLGRAGTQDPTMRALRRYAGVDSADQVLELQHQKRRASDOS  
VNAVREGDVPVIMIGSSTKGAQALAGLTPARAHAPALLBALTLYRHRKPS  
ESLDPRYAVNCIGVCADTDEAKRLRSTQGFALRRGRGLLPVPVDIREHASP  
GELAGIDHTLOYAAGSPETVRKKIDRYVLELTGADBLMAAQIHDAARRRSYELLAS  
LRA"  
complement(1417..2625)  
/gene="CC1958"  
complement(1417..2625)  
/gene="CC1958"  
/note="identified by match to protein family HMM"  
/codon\_start=1  
/transl\_table=11  
/product="glycosyl transferase, group 1 family protein"  
/protein\_id="AAK23933.1"  
/db\_xref="GI:13423419"  
/translation="MTDRPQSLGPGGLQPETRSIGDEPVRVLTDTMLYAPRSGGVR  
VLSSKRWMLASHRPVRRHTLVVPGARDFHDGQVSIYAAPLPGAGVRWPVYKQTM  
ERLIQQPDLIEGDPYTGFLAALRAGDALGVVGFCHTDGLAALHIGMAKRPV  
QKRWAIYRQFQAVAPSRFIEGLIEGVRAIGLPGVDPAIRHPERGGDGLRR  
IGLSPKRLIVFAGRPAREKRLDIVAVVERIGDYLKLFVAGGAPVSPDTCTIDY  
VRDPEVLASILASCAPFVHANNRPFGLVLEAMACGLPYGVSGVAGVESAADAAGV  
LAVASEATRAFADTALFEEDVTAVGEAARRAVRHGMDAVTELCILYDRLTSRA  
FSSASAHAGH"  
complement(2928..3002)  
/gene="trna-val-1"  
complement(2928..3002)  
/gene="trna-val-1"  
/product="trna-val"  
complement(3109..3387)  
/gene="CC1959"  
complement(3109..3387)  
/gene="CC1959"  
/note="identified by match to protein family HMM"  
/codon\_start=1  
/transl\_table=11  
/product="DNA-binding protein HU"  
/protein\_id="AAK23934.1"  
/db\_xref="GI:13423420"  
/translation="MTTKAEVLTATAEKAGINKNKAQKALNFAVYDLSKGDV  
LVGRTFAVTRACATNAPRGTETVNRPAKTAFFVGGGLKSLNS"  
complement(3531..5930)  
/gene="CC1960"  
complement(3531..5930)  
/gene="CC1960"  
/note="identified by match to TIGR protein family HMM  
TIGR00763"  
/codon\_start=1  
/transl\_table=11  
/product="ATP-dependent protease LA"  
/protein\_id="AAK23935.1"  
/db\_xref="GI:13423421"  
/translation="MSELRTLPVLPRLDIVFPHMVVPLEFGGRKSVRALEFYMGRD  
QILLVTOQNSADDDPAPGDIPEVGLATVLOLKLPGDTVAVLVGKARAIVASTDQ  
ESVYEADQIGEVSDGAPAEALSRVAVDFENVYVINKKVPPEALASIPQIAEPK  
LADSIATLHLSVIGDKONLLEIFDVVKRLKLEFVLMGEISVLOVEKKIRSRVKRM  
KTOREYVINEQMKALIORELGPDDARDELIDLEKRIKTKTSKAPRTAESLEKRN  
MSPMSAESTVYVRNLDMLSLTPGKAKTKTKIDVSESGILADADHGLEKVERILEYL

AVOARTNSLKPIILCVGPVGKTSLSKSIATAKGRFVRMSIGVRADEAIRGHR  
TYLGSYDQVSMKAKTNNAPVLDEIDKMGSDYRGPASALVELVDSNOSTED  
HYLEVYDLSQVMEFTTANSLNMPDIDMETIRIDGYTEDEKELIETAKHITIPIKLAK  
DHGLKRAEPIVPPKAIKIDRLIRYTRREGAVSLEREGLAKRTVRLDARKKASITID  
DEBLAKYAGVKATRYGETDEVQVIGVTLAMEFGDILITIAVKRPGKGRMOTGN  
LKDVMEESIANSYVSRALQGIKPPVEKEDVHIVHDGATPKPGSPAGIAMALA  
MYSVLGIPIRKDIAMTGETILRGRTAIGLEKELIALRSQVKTIVLIIQENKDLA  
DVPQTVKGLIEIIPSTVDEVLEKHALTGPLTPVEMWNEAEPIITTSAKKDDGSDAMLT  
H"  
/complement(6325..7587)  
/gene="CC1961"  
/complement(6325..7587)  
/gene="CC1961"  
/note="Identified by match to TIGR protein family HMM  
TIGR00382"  
/codon\_start=1  
/transl\_table=11  
/product="ATP-dependent Clp protease, ATP-binding subunit  
ClpX"  
/protein\_id="AAK23936.1"  
/db\_xref="GI:13423422"  
/translation="MTKRAAGDTRKSTLYCSFCGKSQHEVRKLIAGPTVFIQDEVC  
MDIIEBHKAIEVKSQDVPPTPREICEVLDIVYIGQHAKVLAVAHNHYKRLNHS  
KNDVELAKSNILLVQPTGKTLLAQTLLRIIDVPTMADATTLTEAGYVGEDVENI  
VLKLDADYNVERAQRGIYIDEIDKISRKSNSPSTIRDSGEVQOALLKIMEGT  
ASVPGGRKHPOEFLQVDTNIFLCGAPAGLEKIIISARGAASIGFGAKVTOPE  
ERRTGELRNVEDDLQRFGLIPEFTGLPVVATIEDLEALYKILITPEKNAFVKOY  
ORLPENIGLTETEDALHQAARAKIARTKGARLSIMEGILLETMEFLPTIEGVEE  
VVVNAEYVGRAPLTIYAEKKGASAA"  
7971..8591  
/gene="CC1962"  
7971..8591  
/gene="CC1962"  
/note="Identified by match to protein family HMM"  
/codon\_start=1  
/transl\_table=11  
/product="cIca protein"  
/protein\_id="AAK23937.1"  
/db\_xref="GI:13423423"  
/translation="MTGEMGHEPLIYAFDFDGLTVKDSFNAPLKMWRAGRWSGVLR  
LTPALLAYVFDNRNKGKIAAAYROFLAKGATVAOENDARPAEFAPSLRPDAVAYW  
RGNRAKAKMVIYASFDLIVAPPARGLDLIGTLRCSDDRIIGLDGNNCRAK  
EKVIRLEVFVGPVRLTAAAGDYSGTEMLAIADKGYRIFRGPAA"  
complement(8670..9302)  
/gene="CC1963"  
complement(8670..9302)  
/gene="CC1963"  
/note="Identified by match to TIGR protein family HMM  
TIGR00493"  
/codon\_start=1  
/transl\_table=11  
/product="ATP-dependent Clp protease, proteolytic subunit"  
/protein\_id="AAK23938.1"  
/db\_xref="GI:13423424"  
/translation="MMYDPVSTAMNLVPMVVEQTSRGERAFDIFSRLKRIIIFLTGP  
VEDGMSLICAQLFLESENPKKEIAMYINSPGAVYAGLAITYDMQIKSPVSTCM  
GMAWSMSLLAAGAGORISLPNARIWHPGSGFRGQASDIERHAEDITIKRRRLN  
EIVYKCGRTVEEVERTLDRDHMSADEAKAMGLVHVYSDRAAENGAE"  
complement(9480..10838)  
/gene="CC1964"  
complement(9480..10838)  
/gene="CC1964"  
/note="Identified by match to TIGR protein family HMM  
TIGR00115"  
/codon\_start=1  
/transl\_table=11  
/product="trigger factor"  
/protein\_id="AAK23939.1"  
/db\_xref="GI:13423425"  
/translation="MOIYKSGSLSVPGVTVPASRLATRLERARIEVAPOMWKGF  
RPGKVPAAHRYRKLKMEVIEQALNETTYVLEDNKLIRPACQPELNPSSDMDKYI  
AGGEDLSFDLAVEMPEFEPIPTSTIELVAVYKVSDEYQEADELAKOARTYEPRT  
GKSLAKKDGQDLLIDFVGTIDGVEFAGKAELVYSGQFIQGFEDOLVGAKPGDD

BASE COUNT 1854 a 3619 c 3396 g 2091 t  
ORIGIN

Query Match 92.4%; Score 19.4; DB 1; Length 10960;  
Best Local Similarity 95.2%; Pred. No. 57;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GCGTGGCGAAGACTCGCTGG 21  
|||||||  
Db 4867 GCGTGGCGAAGACTCGCTGG 4847

Search completed: March 23, 2003, 16:10:00  
Job time : 77.4223 secs

VVVKPEPEYQAKDLAKDAEFATKYOEVRAPYDGKADELAKRLGLSDLAUTELL  
KSNLAGRYDSSRRPKIKRALIDVLDTKHDEPLPRMWDAPAGTIWQOVADKARGLP  
PEDAEKTEDLNKDEYKRIARVRRLGLVLAIEIGKNNVYVTTDOELTAINREARQYGA  
QAOQVFMTRQADLOALRAPYIEDKVVDLIFEKATIEKEVSKDELLEDLPEST  
GG"

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 23, 2003, 13:56:37 ; Search time 24.4179 Seconds  
(without alignments)  
1936.775 Million cell updates/sec

Title: US-09-873-546-6

Perfect score: 21

Sequence: 1 gcgtggcgaagaactgcgtg 21

Scoring table:

IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.Geneseq\_101002.\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19.4	92.4	1249	24	ABO72525
2	19.4	92.4	3061	24	ABO72648
3	17.8	84.8	130	24	ABK87227
4	17.8	84.8	400	24	AAI40743
5	17.8	84.8	777	19	AAV19498
6	17.8	84.8	702	23	ABL14159
7	17.8	84.8	1449	24	ABN97193
8	17.8	84.8	1450	22	AAD09849
9	17.8	84.8	1494	16	AAO87729

10	17.8	84.8	1494	16	AAO87730	Human auxillary cy
11	17.8	84.8	1494	16	AAO87731	Human auxillary cy
12	17.8	84.8	1494	16	AAO87732	Human auxillary cy
13	17.8	84.8	1494	17	AAT28395	Human cytochrome p
14	17.8	84.8	1494	17	AAT28396	Human cytochrome p
15	17.8	84.8	1494	17	AAT28397	Human cytochrome p
16	17.8	84.8	1494	17	AAT28398	Human cytochrome p
17	17.8	84.8	1494	17	AAT17388	Human derived cyto
18	17.8	84.8	1494	24	ABO72216	Human CYP2D6 gene
19	17.8	84.8	1545	19	AAV19496	Cytochrome P45011d
20	17.8	84.8	1586	12	AAQ12893	Debrisoquine hydro
21	17.8	84.8	1593	22	AAD09937	Human drug metabol
22	17.8	84.8	2951	23	ABL14158	Drosophila melanog
23	17.8	84.8	3493	23	ABL11094	Drosophila melanog
24	17.8	84.8	6472	24	ABO72215	Human CYP2D6 gene
25	17.8	84.8	6472	24	ABO72364	Human CYP2D6 gene
26	17.8	84.8	9432	24	ABD34213	Human cytochrome p
27	17.4	82.9	922	23	ABL04759	Drosophila melanog
28	17.4	82.9	1471	23	ABL15253	Drosophila melanog
29	17.4	82.9	3350	23	ABL04758	Drosophila melanog
30	17.4	82.9	4389	23	ABL04744	Drosophila melanog
31	17.4	82.9	14112	23	ABL15252	Drosophila melanog
32	16.8	80.0	421	23	AAS82963	DNA encoding novel
33	16.8	80.0	657	24	ABK14232	Human related RAS
34	16.8	80.0	1021	24	AB199215	Mouse ischaemic co
35	16.8	80.0	1078	23	AAS82965	DNA encoding novel
36	16.8	80.0	1188	23	ABL11393	Drosophila melanog
37	16.8	80.0	1629	24	ABO9083	M. capsulatus gene
38	16.8	80.0	2223	22	AAH41606	Human R-Ras, c-Raf
39	16.8	80.0	2223	24	ABL50193	Human R-Ras, c-Raf
40	16.8	80.0	2223	24	ABL50225	Human R-Ras, c-Raf
41	16.8	80.0	2236	17	AAT08522	Oncogene R-Ras mut
42	16.8	80.0	2336	24	ABN95766	Gene #2264 used to
43	16.8	80.0	2813	23	ABL26308	Drosophila melanog
44	16.8	80.0	3188	23	ABL11392	Drosophila melanog
45	16.8	80.0	5686	24	ABK14231	Human related RAS

#### ALIGNMENTS

RESULT 1	ABO72525	standard; CDNA: 1249 BP.
ID	ABO72525	
XX	ABO72525;	
AC	03-SEP-2002	(first entry)
DT	03-SEP-2002	
XX		
DE	Human MDT encoding CDNA seq ID NO 77.	
XX		
XX		
OS	Homo sapiens.	
XX		
PN	WO200240715-A2.	
XX		
PD	23-MAY-2002.	
XX		
PF	06-SEP-2001;	2001WO-US27628.
XX		
PR	06-SEP-2000;	2000US-230505P.
XX		
PR	06-SEP-2000;	2000US-230514P.
XX		
PR	06-SEP-2000;	2000US-230515P.
XX		
PR	06-SEP-2000;	2000US-230517P.
XX		
PR	06-SEP-2000;	2000US-230518P.
XX		
PR	06-SEP-2000;	2000US-230519P.

PR 06-SEP-2000; 2000US-230595P.  
PR 06-SEP-2000; 2000US-230597P.  
PR 06-SEP-2000; 2000US-230598P.  
PR 06-SEP-2000; 2000US-230599P.  
PR 06-SEP-2000; 2000US-230610P.  
PR 06-SEP-2000; 2000US-230655P.  
PR 06-SEP-2000; 2000US-230988P.  
PR 06-SEP-2000; 2000US-230989P.  
PR 07-SEP-2000; 2000US-230951P.  
PR 07-SEP-2000; 2000US-231163P.  
PR 07-SEP-2000; 2000US-231167P.  
PA (INCYTE GENOMICS INC.  
PI Jackson S, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL;  
PI Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE, Dahl CR;  
PI Momiyama MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM;  
PI Gerstlin EH, Peralta CH, David MH, Panzer SR, Flores V, Dafio A;  
PI Marwaha R, Chen AJ, Chang SC, Au AP, Inman RR;  
DR WPI: 2002-527544/56.  
XX P-PSDB; ABP51307.

PT Novel human disease detection and treatment polypeptide, useful in  
PT diagnosis, prevention or treatment of cell proliferative disorders e.g.  
PT arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder  
PT e.g. AIDS

PS Claim 1: Page 350-351. 618pp: English.

XX The invention relates to an isolated human disease detection and  
XX treatment (MDPT) polypeptide (I) selected from a polypeptide having a  
XX sequence selected from 254 sequences (ABP51231-ABP51484) given in the  
XX specification, a naturally occurring polypeptide comprising a sequence  
XX having at least 90% identity to (I) or a biologically active or  
XX immunogenic fragment of (I). (I) is useful for screening a compound for  
XX effectiveness as an agonist or antagonist, for screening a compound that  
XX specifically binds (I) or modulates the activity of (I), and for  
XX preparing a polyclonal or monoclonal antibody by hybridoma technology.  
XX Nucleic acids (II) (ABQ72449-ABQ72700) encoding (I) are useful for  
XX screening a compound for effectiveness in altering expression of a target  
XX polynucleotide comprising. Oligonucleotides and antibodies are useful for  
XX detecting MDPT in a sample or for assessing toxicity of a test compound,  
XX in a diagnostic test for a condition or a disease associated with the  
XX expression of MDPT in a biological sample, for detecting (I) in a sample,  
XX and for purifying (I) from a sample. A composition comprising (I), an  
XX agonist or antagonist is useful for treating a disease or condition  
XX associated with decreased or increased expression of functional MDPT.  
XX (I) or (II) are useful for diagnosing, treating or preventing disorders  
XX associated with aberrant expression of MDPT, where the disorders are  
XX selected from a cell proliferative disorder such as arteriosclerosis,  
XX cirrhosis, hepatitis, psoriasis, and cancer and an  
XX autoimmune/inflammatory disorder such as AIDS, Addison's disease,  
XX allergy, anaemia, asthma, atherosclerosis, gout, multiple sclerosis or  
XX rheumatoid arthritis. (II) are useful for creating knockin humanised  
XX animals or transgenic animals to model human diseases, in somatic or  
XX germline gene therapy, to generate a transcript image of a tissue or cell  
XX type, for detecting differences in the chromosomal location due to  
XX translocation or inversion among normal, carrier or affected individuals  
XX and as hybridisation probes for mapping naturally occurring genomic  
XX sequences.

SQ Sequence 1249 BP; 230 A; 436 C; 355 G; 228 T; 0 other;

Query Match 92.4%; Score 19.4; DB 24; Length 1249;

Best Local Similarity 95.2%; Pred. No. 4.4;

Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCGTGGCAGAGACTGCTGG 21

DB 100 GCGTGGCAGAGACTGCTGG 120

RESULT 2  
ABQ72648  
ID ABQ72648 standard; cDNA; 3061 BP.  
XX  
AC ABQ72648;  
XX  
DT 03-SEP-2002 (first entry)  
XX  
DE Human MDPT encoding cDNA SEQ ID NO 200.  
XX  
KW Human; MDPT; disease detection and treatment molecule polynucleotide;  
KW proliferative disorder; hepatitis; psoriasis; cancer; AIDS;  
KW autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;  
KW rheumatoid arthritis; transgenic; gene therapy; antiarteriosclerosis;  
KW hepatocytic; antiinflammatory; antiproliferative; cytostatic; anti-HIV;  
KW antiangiogenic; antianemic; antistatic; antiatherosclerotic; antiout;  
KW neuroprotective; antineumatic; antirheumatic; gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN MO200240715-A2.  
XX  
PD 23-MAY-2002.  
XX  
PF 06-SEP-2001; 2001WO-US27628.  
XX  
PR 06-SEP-2000; 2000US-230505P.  
PR 06-SEP-2000; 2000US-230514P.  
PR 06-SEP-2000; 2000US-230515P.  
PR 06-SEP-2000; 2000US-230517P.  
PR 06-SEP-2000; 2000US-230518P.  
PR 06-SEP-2000; 2000US-230519P.  
PR 06-SEP-2000; 2000US-230595P.  
PR 06-SEP-2000; 2000US-230597P.  
PR 06-SEP-2000; 2000US-230598P.  
PR 06-SEP-2000; 2000US-230599P.  
PR 06-SEP-2000; 2000US-230610P.  
PR 06-SEP-2000; 2000US-230655P.  
PR 06-SEP-2000; 2000US-230988P.  
PR 06-SEP-2000; 2000US-230989P.  
PR 07-SEP-2000; 2000US-230951P.  
PR 07-SEP-2000; 2000US-231163P.  
PR 07-SEP-2000; 2000US-231167P.  
XX  
PA (INCYTE GENOMICS INC.  
PI Jackson S, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL;  
PI Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE, Dahl CR;  
PI Momiyama MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM;  
PI Gerstlin EH, Peralta CH, David MH, Panzer SR, Flores V, Dafio A;  
PI Marwaha R, Chen AJ, Chang SC, Au AP, Inman RR;  
DR WPI: 2002-527544/56.  
XX P-PSDB; ABP51431.

PT Novel human disease detection and treatment polypeptide, useful in  
PT diagnosis, prevention or treatment of cell proliferative disorders e.g.  
PT arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder  
PT e.g. AIDS

PS Claim 1: Page 414; 618pp: English.

XX The invention relates to an isolated human disease detection and  
XX treatment (MDPT) polypeptide (I) selected from a polypeptide having a  
XX sequence selected from 254 sequences (ABP51231-ABP51484) given in the  
XX specification, a naturally occurring polypeptide comprising a sequence  
XX having at least 90% identity to (I) or a biologically active or  
XX immunogenic fragment of (I). (I) is useful for screening a compound for  
XX effectiveness as an agonist or antagonist, for screening a compound that  
XX specifically binds (I) or modulates the activity of (I), and for  
XX preparing a polyclonal or monoclonal antibody by hybridoma technology.  
XX Nucleic acids (II) (ABQ72449-ABQ72700) encoding (I) are useful for  
XX screening a compound for effectiveness in altering expression of a target



CC polynucleotide comprising, oligonucleotides and antibodies are useful for  
 CC detecting MDR in a sample or for assessing toxicity of a test compound,  
 CC in a diagnostic test for a condition or a disease associated with the  
 CC expression of MDR in a biological sample, for detecting (I) in a sample,  
 CC and for purifying (I) from a sample. A composition comprising (I), an  
 CC agonist or antagonist is useful for treating a disease or condition  
 CC associated with decreased or increased expression of functional MDR.  
 CC (I) or (II) are useful for diagnosing, treating or preventing disorders  
 CC associated with aberrant expression of MDR, where the disorders are  
 CC selected from a cell proliferative disorder such as arteriosclerosis,  
 CC cirrhosis, hepatitis, psoriasis, and cancer and an  
 CC autoimmune/inflammatory disorder such as AIDS, Addison's disease,  
 CC allergy, anemia, asthma, atherosclerosis, gout, multiple sclerosis or  
 CC rheumatoid arthritis. (II) are useful for creating knockin humanised  
 CC animals or transgenic animals to model human diseases, in somatic or  
 CC germline gene therapy, to generate a transcript image of a tissue or cell  
 CC type, for detecting differences in the chromosomal location due to  
 CC translocation or inversion among normal, carrier or affected individuals  
 CC and as hybridisation probes for mapping naturally occurring genomic  
 CC sequences.

XX  
 SQ Sequence 3061 BP; 606 A; 840 C; 927 G; 688 T; 0 other;

Query Match 92.4%; Score 19.4; DB 24; Length 3061;  
 Best Local Similarity 95.2%; Pred. No. 4.9;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GCGTGGCGAAGACTCGCTGG 21  
 |||||||||||  
 Db 100 GCGTGGCGAAGACTCGCTGG 120

RESULT 3  
 ABR87227  
 ID ABR87227 standard; DNA; 130 BP.  
 AC ABR87227;  
 XX  
 DT 24-SEP-2002 (first entry)  
 XX  
 DE Partial human cytochrome 2D6 gene sequence.

XX  
 KW Restriction endonuclease recognition sequence; RERS; drug resistance;  
 KW human genetic disease; pathogenic microorganism; bladder carcinoma;  
 KW sickle-cell anaemia; thalassemia; Alzheimer's disease; phenylketonuria;  
 KW galactosaemia; Wilson's disease; diabetes insipidus; neurofibromatosis;  
 KW familial hypercholesterolemia; genetic analysis; hereditary disease;  
 KW tumour diagnosis; disease predisposition; forensic; paternity;  
 KW crop cultivation; animal breeding; expression profiling;  
 KW infectious organism; plant; food safety; cytochrome 2D6;  
 KW single nucleotide polymorphism; SNP; gene; ds.

XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT variation replace (65..67,GC)  
 FT /tag= a  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT /note= "SNP"

XX  
 PN WO200246447-A2.  
 PD 13-JUN-2002.  
 PF 01-OCT-2001; 2001WO-US30742.  
 XX  
 PR 02-OCT-2000; 2000US-237409P.  
 PR 10-NOV-2000; 2000US-247166P.  
 PR 10-NOV-2000; 2000US-247167P.  
 PR 10-NOV-2000; 2000US-247172P.  
 PR 10-NOV-2000; 2000US-247173P.  
 PR 10-NOV-2000; 2000US-247275P.  
 PR 24-JAN-2001; 2001US-263971P.

PR 15-FEB-2001; 2001US-269244P.  
 PR 21-JUN-2001; 2001US-300319P.  
 PR 21-JUN-2001; 2001US-300350P.  
 PR 27-JUN-2001; 2001US-301394P.  
 XX  
 PA (KECK-) KECK GRADUATE INST.  
 XX  
 PI Van Ness J, Galas DJ, Garrison LK;  
 XX  
 DR WPI: 2002-527924/56.  
 XX  
 PT Identifying single nucleotide polymorphisms at defined positions in  
 PT target nucleic acids, by utilising oligonucleotide primers that contain  
 PT a part of an interrupted restriction endonuclease recognition sequence

XX  
 PS Example: Page 54; 82pp; English.

XX  
 CC The present invention relates to a new method of identifying a nucleotide  
 CC at a defined position in a single stranded target nucleic acid. The  
 CC method of the invention utilises a pair of oligonucleotide primers  
 CC comprising a nucleotide sequence that is complementary to a nucleotide  
 CC sequence of the target nucleic acid or complementary to a nucleotide  
 CC sequence of the complement of the target nucleic acid and further  
 CC comprising part of an interrupted restriction endonuclease recognition  
 CC sequence (RERS). The invention is useful for identifying a nucleotide at  
 CC a defined position in a single-stranded target nucleic acid. The defined  
 CC position may be polymorphic or associated with a disease, especially a  
 CC human genetic disease, or drug resistance of a pathogenic microorganism.  
 CC The defined position is associated with a disease, including a human  
 CC genetic disease e.g. bladder carcinoma, sickle-cell anaemia,  
 CC thalassemias, Alzheimer's disease, phenylketonuria, galactosaemia,  
 CC Wilson's disease, diabetes insipidus, familial hypercholesterolemia, or  
 CC neurofibromatosis. The invention finds applications in genetic analysis  
 CC for hereditary disease, tumour diagnosis, disease predisposition,  
 CC forensics or paternity, crop cultivation and animal breeding, expression  
 CC profiling of cell function and/or disease marker genes, and  
 CC identification and/or characterisation of infectious organisms that  
 CC cause infectious disease in plants or animals and/or that are related  
 CC to food safety. The method is useful for determining gene variations in  
 CC T-cell receptor genes encoding variable, antigen-specific regions that  
 CC are involved in the recognition of various foreign antigens. The present  
 CC nucleic acid sequence represents the partial human cytochrome 2D6 gene  
 CC sequence that was used in the method of the invention.

XX  
 SQ Sequence 130 BP; 19 A; 40 C; 46 G; 25 T; 0 other;

Query Match 84.8%; Score 17.8; DB 24; Length 130;  
 Best Local Similarity 90.5%; Pred. No. 22;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GCGTGGCGAAGACTCGCTGG 21  
 |||||||||||  
 Db 40 GCGTGGCGAAGACTCGCTGG 60

RESULT 4  
 AAL40743  
 ID AAL40743 standard; DNA; 400 BP.  
 AC AAL40743;  
 XX  
 DT 25-SEP-2002 (first entry)  
 XX  
 DE DNA sequence of amplifier containing CYP450-2D6-G1749C SNP.  
 XX  
 KW Variation site: analysing; point mutation; detecting pathogen; SNP;  
 KW single nucleotide polymorphisms; paternity dispute; prenatal testing;  
 KW forensic analysis; CYP450-2D6-G1749C; ds.  
 XX  
 OS unidentified.  
 XX  
 PN WO200194546-A2.



CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 702 BP; 185 A; 187 C; 220 G; 110 T; 0 other;

Query Match 84.8%; Score 17.8; DB 23; Length 702;

Best Local Similarity 90.5%; Pred. No. 26;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

YY 1 GCCTGGCAGAACTCGCTGG 21  
|||

DB 83 GCCTGGCAGAACTCGCTGG 103

RESULT 7

ABN97193  
ID ABN97193 standard; DNA: 1449 BP.

AC ABN97193;

DT 13-AUG-2002 (first entry)

DE Gene #3691 used to diagnose liver cancer.

KM Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;

KM metastatic liver tumor; cytostatic; expression profile; disease state;

KM disease progression; drug toxicity; drug efficacy; drug metabolism.

OS Homo sapiens.

PN WO200229103-A2.

PD 11-APR-2002.

PF 02-OCT-2001; 2001WO-US30589.

PR 02-OCT-2000; 2000US-237054P.

PA (GENE-) GENE LOGIC INC.

PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;

DR WPI; 2002-426119/45.

PT Diagnosing and detecting the progression of liver cancer,

PT hepatocellular carcinoma or metastatic liver tumor in a patient,

PT involves detecting the level of expression of two or more genes in a

XX Claim 1; SEQ ID NO 3691; 298bp; English.

XX The invention relates to a novel method for diagnosing and detecting the

XX progression of liver cancer, hepatocellular carcinoma or metastatic liver

XX tumor in a patient, and differentiating metastatic liver cancer from

XX hepatocellular carcinoma in a patient, involving detecting the level of

XX expression of two or more genes represented in ABN93503-ABN97455 in a

XX tissue sample. The method of the invention has hepatotropic, and

XX cytostatic activity. The method is useful for diagnosing and detecting

XX the progression of liver cancer, hepatocellular carcinoma and metastatic

XX liver carcinoma in a patient. The method is useful for identifying

XX expression profiles which serve as useful diagnostic markers as well as

XX markers that can be used to monitor disease states, disease progression,

XX drug toxicity, drug efficacy and drug metabolism.

XX Note: The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 1449 BP; 256 A; 465 C; 439 G; 289 T; 0 other;

XX Query Match 84.8%; Score 17.8; DB 24; Length 1449;

XX Best Local Similarity 90.5%; Pred. No. 29;

XX Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

YY 1 GCCTGGCAGAACTCGCTGG 21  
|||

DB 131 GCCTGGCAGAACTCGCTGG 151

RESULT 8

AA09849  
ID AA09849 standard; DNA: 1450 BP.

AC AA09849;

DT 12-SEP-2001 (first entry)

DE Human CYP2D6 gene.

KM Polymorphism; amplification; CYP2D6; cytochrome P450; CYP; human;

KM drug metabolism; psychiatric disorder; cardiovascular disorder; ds.

OS Homo sapiens.

PN WO200149883-A2.

PD 12-JUL-2001.

PF 22-DEC-2000; 2000WO-US35186.

PR 30-DEC-1999; 99US-0173699.

PA (ABBO ) ABBOTT LAB.

PI Katz DA, Gentile-Davey MC, Cornwell MJ, Huff JB;

DR WPI; 2001-441898/47.

PT Detecting a mutation in target nucleic acid sequence in test sample, by

PT amplifying target and standard nucleic acid sequence using primers,

PT hybridizing probes to the products to form hybrids, and detecting

XX Example 1; Page 30; 35pp; English.

XX The invention relates to a method for detecting polymorphism in a target

XX nucleic acid sequence using amplification technique. The method involves

XX amplifying the target sequence and a standard nucleic acid sequence

XX using primers to form amplification products, hybridizing a first

XX labelled probe to the target sequence amplification product and a

XX second labelled probe to the standard sequence amplification product,

XX detecting the signals from the first and the second probe, and comparing

XX the signals to determine the polymorphism. The method is useful for

XX detecting polymorphism in various nucleic acid sequences e.g. CYP2D6 gene

XX which is a member of cytochrome P450 (CYP) gene family. CYP2D6 plays a

XX role in the metabolism of several drugs, including those used for

XX treating psychiatric and cardiovascular disorders. Polymorphism in

XX the CYP2D6 gene have varying effect on an individual's ability to

XX metabolise drugs. The method is suitable for detecting amplification

XX products from multiple and different types of polymorphisms on a single

XX automated platform. The present sequence is human CYP2D6 gene.

XX Sequence 1450 BP; 270 A; 395 C; 521 G; 264 T; 0 other;

XX Query Match 84.8%; Score 17.8; DB 22; Length 1450;

XX Best Local Similarity 90.5%; Pred. No. 29;

XX Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

XX YY 1 GCCTGGCAGAACTCGCTGG 21  
|||

XX DB 150 GCCTGGCAGAACTCGCTGG 170

XX RESULT 9

XX AA087729

XX ID AA087729 standard; cDNA: 1494 BP.

```
XX AC AAQ87729;
XX XX
XX 15-NOV-1995 (first entry)
XX DE Human auxillary cytochrome P450 species 2D6 coding region.
XX DE
XX DE Human cytochrome P450: amplification; PCR; primer; expression vector;
XX DE yeast NADPH-P450 reductase; safety; fusion protein; metabolite;
XX DE carcinogen; mutagen; liver metabolism; ds.
XX OS Homo sapiens.
XX PN EP644267-A.
XX PD 22-MAR-1995.
XX PF 20-JUL-1994; 94EP-0111298.
XX PR 21-JUL-1993; 93JP-0180246.
XX PR 20-JUL-1993; 93JP-0201120.
XX PR 30-JUL-1993; 93JP-0208279.
XX PA (HAYASHI) HAYASHI K.
XX PA (SUMO) SUMITOMO CHEM CO LTD.
XX PI Hayashi K, Kaneko H, Komai K, Nakatsuka I, Sakaki T;
XX PI Yabusaki Y;
XX DR WPI: 1995-116991/16.
XX DR P-PSDB: AAR72375.
XX PT Evaluation of safety of a chemical cpd. - using recombinant yeast
XX PT expressing human cytochrome p450 and a yeast NADPH-P450 reductase
XX PS Examples; Page 82-84, 124pp; English.
XX XX
XX The nucleotide sequence of the cDNA coding region for the human
XX CC auxillary cytochrome P450 species 2C6. The gene encodes a protein of 497
XX CC amino acids. The cDNA was amplified by PCR using the primers AAQ87763-6.
XX CC The product was cloned into the yeast expression vectors pAAH5N or pAHR
XX CC to produce the vectors p2D6 for the expression of the cytochrome P450
XX CC alone or p2D6R for co-expression with the yeast NADPH-P450 reductase.
XX CC The vectors are used in a method for evaluating the safety of a chemical
XX CC compound by reacting the chemical compound with recombinantly produced
XX CC human cytochrome P450 molecular species 1A2 (AAQ87714), 2C9 (AAQ87715),
XX CC 2E1 (AAQ87716), or 3A4 (AAQ87717) or their auxillary species and
XX CC variants (AAQ87718-32), and yeast NADPH-P450 reductase, either as a fused
XX CC protein or in cell extracts, and analysing the resulting metabolite to
XX CC assess the safety of the chemical compound. The method is useful for
XX CC determining whether the chemical compound, or its metabolite, will be
XX CC converted into a carcinogenic or mutagenic form through metabolism in the
XX CC liver.
XX CC
XX Sequence 1494 BP; 248 A; 508 C; 446 G; 292 T; 0 other;
XX
XX Query Match 84.8%; Score 17.8; DB 16; Length 1494;
XX Best Local Similarity 90.5%; Pred. No. 29;
XX Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 GCGTGGCAGAACTCGCTGG 21
OY 11111111111111111111
DB 428 GCGTGGCAGAACTCGCTGG 448
RESULT 10
ID AAQ87730 standard; cDNA: 1494 BP.
XX AAQ87730;
XX AC
XX XX
XX DT 15-NOV-1995 (first entry)
XX XX
```

```
DE DE Human auxillary cytochrome P450 species 2D6 variant 1 coding region.
XX XX
XX XX Human cytochrome P450: amplification; PCR; primer; expression vector;
XX KW yeast NADPH-P450 reductase; safety; fusion protein; metabolite;
XX KW carcinogen; mutagen; liver metabolism; ds.
XX XX
XX OS Homo sapiens.
XX PN EP644267-A.
XX PD 22-MAR-1995.
XX PF 20-JUL-1994; 94EP-0111298.
XX PR 21-JUL-1993; 93JP-0180246.
XX PR 20-JUL-1993; 93JP-0201120.
XX PR 30-JUL-1993; 93JP-0208279.
XX PA (HAYASHI) HAYASHI K.
XX PA (SUMO) SUMITOMO CHEM CO LTD.
XX PI Hayashi K, Kaneko H, Komai K, Nakatsuka I, Sakaki T;
XX PI Yabusaki Y;
XX DR WPI: 1995-116991/16.
XX DR P-PSDB: AAR72376.
XX PT Evaluation of safety of a chemical cpd. - using recombinant yeast
XX PT expressing human cytochrome p450 and a yeast NADPH-P450 reductase
XX PS Examples; Page 87-89; 124pp; English.
XX XX
XX The nucleotide sequence of the cDNA coding region for the human
XX CC auxillary cytochrome P450 species 2D6 variant 1. The gene contains a
XX CC change at base 886 from T to C as compared to the wild type sequence
XX CC (AAQ87729). This changes the amino acid residue 296 from Cys to Arg. The
XX CC cDNA was amplified by PCR using the primers AAQ87763-6. The product was
XX CC cloned into the yeast expression vectors pAAH5N or pAHR to produce the
XX CC vectors p2D6 variant 1 for the expression of the cytochrome P450 alone
XX CC or p2D6R variant 1 for co-expression with the yeast NADPH-P450
XX CC reductase. The vectors are used in a method for evaluating the safety of
XX CC a chemical compound by reacting the chemical compound with recombinantly
XX CC produced human cytochrome P450 molecular species 1A2 (AAQ87714), 2C9
XX CC (AAQ87715), 2E1 (AAQ87716), or 3A4 (AAQ87717) or their auxillary species
XX CC and variants (AAQ87718-32), and yeast NADPH-P450 reductase, either as a
XX CC fused protein or in cell extracts, and analysing the resulting metabolite
XX CC to assess the safety of the chemical compound. The method is useful for
XX CC determining whether the chemical compound, or its metabolite, will be
XX CC converted into a carcinogenic or mutagenic form through metabolism in the
XX CC liver.
XX CC
XX Sequence 1494 BP; 248 A; 509 C; 446 G; 291 T; 0 other;
XX
XX Query Match 84.8%; Score 17.8; DB 16; Length 1494;
XX Best Local Similarity 90.5%; Pred. No. 29;
XX Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 GCGTGGCAGAACTCGCTGG 21
OY 11111111111111111111
DB 428 GCGTGGCAGAACTCGCTGG 448
RESULT 11
ID AAQ87731 standard; cDNA: 1494 BP.
XX AAQ87731;
XX AC
```

```

XX 15-NOV-1995 (first entry)
DE Human auxillary cytochrome P450 species 2D6 variant 2 coding region.
XX
XX Human cytochrome P450; amplification; PCR; primer; expression vector;
KW yeast NADPH-P450 reductase; safety; fusion protein; metabolite;
KW carcinogen; mutagen; liver metabolism; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FH variation 886
FT /tag= a
FT /note= "T to C change in variant 2 changes amino
FT 1457 acid from Cys to Arg"
FT variation
FT /tag= b
FT /note= "C to G change in variant 2 changes amino
FT acid from Thr to Ser"
XX
XX EP644267-A.
XX
XX 22-MAR-1995.
XX
XX 20-JUL-1994; 94EP-0111298.
XX
XX 21-JUL-1993; 93JP-0180246.
XX 20-JUL-1993; 93JP-0201120.
XX 30-JUL-1993; 93JP-0208279.
XX
XX (HAYASHI K.
XX (SUMO ) SUMITOMO CHEM CO LTD.
XX
XX Hayashi K, Kaneko H, Komai K, Nakatsuka I, Sakaki T;
XX Yabusaki Y;
XX WPI: 1995-116991/16.
XX P-PSDB; AAR72377.
XX
XX Evaluation of safety of a chemical cpd. - using recombinant yeast
XX expressing human cytochrome p450 and a yeast NADPH-P450 reductase
XX
XX Examples; Page 91-93; 124pp; English.
XX
XX The nucleotide sequence of the cDNA coding region for the human
XX auxillary cytochrome P450 species 2D6 variant 2. The gene contains
XX variations at bases 886: T to C and 1457: C to G as compared to the wild
XX type sequence (AA087729). These changes the amino acid residues 296: Cys
XX to Arg and 486: Thr to Ser. The cDNA was amplified by PCR using the
XX primers AA087763-6. The product was cloned into the yeast expression
XX vectors pAAH5N or pAHR to produce the vectors p2D6 variant 2 for the
XX expression of the cytochrome P450 alone or p2D6R variant 2 for
XX co-expression with the yeast NADPH-P450 reductase.
XX The vectors are used in a method for evaluating the safety of a chemical
XX compound by reacting the chemical compound with recombinantly produced
XX human cytochrome P450 molecular species 1A2 (AA087714), 2C9 (AA087715),
XX 2E1 (AA087716), or 3A4 (AA087717) or their auxillary species and
XX variants (AA087718-32), and yeast NADPH-P450 reductase, either as a fused
XX protein or in cell extracts, and analysing the resulting metabolite to
XX assess the safety of the chemical compound. The method is useful for
XX determining whether the chemical compound, or its metabolite, will be
XX converted into a carcinogenic or mutagenic form through metabolism in the
XX liver.
XX
XX Sequence 1494 BP; 248 A; 508 C; 447 G; 291 T; 0 other;
XX
Query Match 84.8%; Score 17.8; DB 16; Length 1494;
Best Local Similarity 90.5%; Pred. No. 29;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 GCCTGGCCAGAACTCGCTGG 21
II IIIIIIIIIII IIIIIII

```

```

DB 428 GCCTGGCCAGAACTCGCTGG 448
RESULT 12
AA087732
ID AA087732 standard; cDNA; 1494 BP.
XX
XX AA087732;
XX
XX 15-NOV-1995 (first entry)
XX
XX Human auxillary cytochrome P450 species 2D6 variant 3 coding region.
XX
XX Human cytochrome P450; amplification; PCR; primer; expression vector;
KW yeast NADPH-P450 reductase; safety; fusion protein; metabolite;
KW carcinogen; mutagen; liver metabolism; ds.
XX
XX Homo sapiens.
XX
FH Key Location/Qualifiers
FH variation 1457
FT /tag= a
FT /note= "C to G change in variant 1 changes amino
FT acid from Thr to Ser"
XX
XX EP644267-A.
XX
XX 22-MAR-1995.
XX
XX 20-JUL-1994; 94EP-0111298.
XX
XX 21-JUL-1993; 93JP-0180246.
XX 20-JUL-1993; 93JP-0201120.
XX 30-JUL-1993; 93JP-0208279.
XX
XX (HAYASHI K.
XX (SUMO ) SUMITOMO CHEM CO LTD.
XX
XX Hayashi K, Kaneko H, Komai K, Nakatsuka I, Sakaki T;
XX Yabusaki Y;
XX WPI: 1995-116991/16.
XX P-PSDB; AAR72378.
XX
XX Evaluation of safety of a chemical cpd. - using recombinant yeast
XX expressing human cytochrome p450 and a yeast NADPH-P450 reductase
XX
XX Examples; Page 95-97; 124pp; English.
XX
XX The nucleotide sequence of the cDNA coding region for the human
XX auxillary cytochrome P450 species 2D6 variant 3. The gene contains a
XX change at base 1457 from C to G as compared to the wild type sequence
XX (AA087729). This changes the amino acid residue 236 from Thr to Ser. The
XX cDNA was amplified by PCR using the primers AA087763-6. The product was
XX cloned into the yeast expression vectors pAAH5N or pAHR to produce the
XX vectors p2D6 variant 3 for the expression of the cytochrome P450 alone
XX or p2D6R variant 3 for co-expression with the yeast NADPH-P450
XX reductase. The vectors are used in a method for evaluating the safety of
XX a chemical compound by reacting the chemical compound with recombinantly
XX produced human cytochrome P450 molecular species 1A2 (AA087714), 2C9
XX (AA087715), 2E1 (AA087716), or 3A4 (AA087717) or their auxillary species
XX and variants (AA087718-32), and yeast NADPH-P450 reductase, either as a
XX fused protein or in cell extracts, and analysing the resulting metabolite
XX to assess the safety of the chemical compound. The method is useful for
XX determining whether the chemical compound, or its metabolite, will be
XX converted into a carcinogenic or mutagenic form through metabolism in the
XX liver.
XX
XX Sequence 1494 BP; 248 A; 507 C; 447 G; 292 T; 0 other;
XX
Query Match 84.8%; Score 17.8; DB 16; Length 1494;
Best Local Similarity 90.5%; Pred. No. 29;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```





**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.4\_P5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: March 23, 2003, 15:14:42 ; Search time 5.54252 Seconds  
(without alignments)  
2938.192 Million cell updates/sec

Title: US-09-873-546-6  
Perfect score: 21  
Sequence: 1 gcgtggcagaactcgtcg 21

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 53826 seqs, 38737923 residues  
Total number of hits satisfying chosen parameters: 1077652

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_NA:\*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCF\_NEM\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEM\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEM\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCFUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEM\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_NEM\_PUB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US10\_NEM\_PUB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US60\_NEM\_PUB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	17.8	84.8	1449	10	US-09-880-107-3688 Sequence 3688, Ap
2	17.8	84.8	1450	10	US-09-747-538-1 Sequence 1, Appli.
3	17.8	84.8	9432	9	US-09-942-310-1 Sequence 44, Appli
4	16.8	80.0	1157	9	US-10-108-605-44 Sequence 1, Appli
5	16.8	80.0	68750	9	US-10-014-717-1 Sequence 10636, A
6	16.4	78.1	256	10	US-09-960-352-10636 Sequence 2434, Ap
7	16.4	78.1	348	10	US-09-960-352-2434 Sequence 10478, A
8	16.4	78.1	422	10	US-09-960-352-10478 Sequence 13143, A
9	16.2	77.1	385	10	US-09-960-352-13143 Sequence 4246, Ap
10	16.2	77.1	398	10	US-09-960-352-4246 Sequence 4253, Ap
11	16.2	77.1	452	10	US-09-960-352-4253 Sequence 769, App
12	16.2	77.1	454	10	US-09-764-864-769 Sequence 5830, Ap
13	16.2	77.1	498	10	US-09-960-352-5830 Sequence 354, App
14	16.2	77.1	1443	9	US-09-962-832-224 Sequence 224, App
15	16.2	77.1	1511	9	US-10-078-650-11 Sequence 11, Appli
16	16.2	77.1	1964	10	US-09-925-297-367 Sequence 367, App
17	16.2	77.1	2746	10	US-09-917-800A-1463 Sequence 1463, Ap
18	16.2	77.1	3641	9	US-10-078-650-1 Sequence 1, Appli
19	16.2	77.1	3641	9	US-10-078-650-1 Sequence 1, Appli

20	15.8	75.2	185	10	US-09-294-093B-5564 Sequence 5564, Ap
21	15.8	75.2	295	10	US-09-294-093B-1650 Sequence 1650, Ap
22	15.8	75.2	348	9	US-09-960-352-1110 Sequence 1110, Ap
23	15.8	75.2	1749	9	US-09-809-391-54 Sequence 54, Appli
24	15.8	75.2	8439	10	US-09-764-877-3959 Sequence 3959, Ap
25	15.8	75.2	9875	10	US-09-764-877-3960 Sequence 3960, Ap
26	15.8	75.2	31412	9	US-10-109-551-3 Sequence 3, Appli
27	15.8	75.2	78056	9	US-10-109-551-1 Sequence 1, Appli
28	15.4	73.3	48	10	US-09-880-732-56 Sequence 56, Appli
29	15.4	73.3	47	10	US-09-880-732-55 Sequence 55, Appli
30	15.4	73.3	4059	9	US-09-784-554B-1 Sequence 1, Appli
31	15.2	72.4	473	10	US-09-833-790-372 Sequence 372, App
32	15.2	72.4	570	12	US-10-104-484-1 Sequence 1, Appli
33	15.2	72.4	570	12	US-10-104-484-3 Sequence 3, Appli
34	15.2	72.4	751	10	US-09-770-149-6 Sequence 6, Appli
35	15.2	72.4	1387	10	US-09-925-302-217 Sequence 217, App
36	15.2	72.4	1832	10	US-09-822-849A-407 Sequence 407, App
37	15.2	72.4	2136	9	US-09-938-842A-1480 Sequence 1480, Ap
38	15.2	72.4	2508	9	US-09-738-626-2305 Sequence 2305, Ap
39	15.2	72.4	2760	10	US-09-934-868-51 Sequence 51, Appli
40	15.2	72.4	3750	9	US-09-964-868-32 Sequence 32, Appli
41	15.2	72.4	6254	9	US-09-764-868-1495 Sequence 1495, Ap
42	15.2	72.4	6457	10	US-09-880-107-3389 Sequence 3389, Ap
43	15.2	72.4	6491	10	US-09-954-456-2118 Sequence 2118, Ap
44	14.8	70.5	311	10	US-09-864-761-27685 Sequence 27685, A
45	14.8	70.5	412	10	US-09-960-352-14532 Sequence 14532, A

ALIGNMENTS

RESULT 1  
US-09-880-107-3688  
Sequence 3688 Application US/09880107  
Patent No. US20020142981A1  
GENERAL INFORMATION:  
APPLICANT: Horne, Darci T.  
APPLICANT: Vockley, Joseph G.  
APPLICANT: Scherf, Uwe  
APPLICANT: Gene Logic, Inc.  
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer  
FILE REFERENCE: 44921-5028-WO  
CURRENT APPLICATION NUMBER: US/09/880.107  
CURRENT FILING DATE: 2001-06-14  
PRIOR APPLICATION NUMBER: US 60/211.379  
PRIOR FILING DATE: 2000-06-14  
PRIOR APPLICATION NUMBER: US 60/237.054  
PRIOR FILING DATE: 2000-10-02  
NUMBER OF SEQ ID NOS: 3950  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 3688  
LENGTH: 1449  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: Genbank Accession No. US20020142981A1 X07618  
US-09-880-107-3688  
Query Match 84.8%; Score 17.8; DB 10; Length 1449;  
Best Local Similarity 90.5%; Pred. No. 9.2;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 GCGTGGCAGACACGCTGG 21  
DB 131 GCGTGGCAGACACGCTGG 151  
RESULT 2  
US-09-747-538-1  
Sequence 1, Application US/09747538  
Patent No. US20020102549A1  
GENERAL INFORMATION:  
APPLICANT: Abbott Laboratories

```

; APPLICANT: Katz, David A.
; APPLICANT: Gentile-Davey, Maria C.
; APPLICANT: Cornwell, Michael C.
; APPLICANT: Huff, Jeffrey B.
; APPLICANT: Yu, Hong
; TITLE OF INVENTION: AMPLIFICATION BASED POLYMORPHISM
; FILE REFERENCE: 6652.US.01
; CURRENT APPLICATION NUMBER: US/09/747,538
; CURRENT FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1450
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-747-538-1

Query Match
Best Local Similarity 84.8%; Score 17.8; DB 10; Length 1450;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCCTGGGCAAGACTGCTGG 21
Db 150 GCCTGGGCAAGACTGCTGG 170

RESULT 3
US-09-942-310-1
; Sequence 1, Application US/09942310
; Publication No. US20030044797A1
; GENERAL INFORMATION:
; APPLICANT: Ristinger, Carl
; APPLICANT: Andersson, Maria K.
; APPLICANT: Lewander, Tommy
; APPLICANT: Olafsson, Erik
; TITLE OF INVENTION: Detection of CYP2D6 Polymorphisms
; FILE REFERENCE: G6119.1US
; CURRENT APPLICATION NUMBER: US/09/942,310
; CURRENT FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: GB 0021286.0
; PRIOR FILING DATE: 2000-08-30
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 9432
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-942-310-1

Query Match
Best Local Similarity 84.8%; Score 17.8; DB 9; Length 9432;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCCTGGGCAAGACTGCTGG 21
Db 3300 GCCTGGGCAAGACTGCTGG 3320

RESULT 4
US-10-108-605-44
; Sequence 44, Application US/10108605
; Patent No. US20020160934A1
; GENERAL INFORMATION:
; APPLICANT: Brocatus, Julie
; APPLICANT: Stam, Lynn
; APPLICANT: Bachmann, Jane
; APPLICANT: Kamdar, Kim
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCODE
; FILE REFERENCE: 31133B
; CURRENT APPLICATION NUMBER: US/10/108,605
; CURRENT FILING DATE: 2002-03-27
```

```

; PRIOR APPLICATION NUMBER: US 09/761,142
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/176,418
; PRIOR FILING DATE: 2000-01-14
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 44
; LENGTH: 1157
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-10-108-605-44

Query Match
Best Local Similarity 80.0%; Score 16.8; DB 9; Length 1157;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCCTGGGCAAGACTGCTG 20
Db 356 GTGTGGGCAAGACTGCTG 375

RESULT 5
US-10-014-717-1
; Sequence 1, Application US/10014717
; Publication No. US20020192778A1
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/10/014,717
; CURRENT FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US/09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-10-014-717-1

Query Match
Best Local Similarity 80.0%; Score 16.8; DB 9; Length 68750;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCCTGGGCAAGACTGCTG 20
Db 21493 GCCTGGGCAAGACTGCTG 21512

RESULT 6
US-09-960-352-10636/c
; Sequence 10636, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathalagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 10636
; LENGTH: 256
; TYPE: DNA
```

ORGANISM: Bos taurus  
OTHER INFORMATION: Clone ID: 46-LIB3058-004-Q1-E1-D10  
US-09-960-352-10636

Query Match 78.1%; Score 16.4; DB 10; Length 256;  
Best Local Similarity 94.4%; Pred. No. 39;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 TGGCAGAACTCGCTGG 21  
DB 142 TGACGAGAACTCGCTGG 125

## RESULT 7

US-09-960-352-2434  
Sequence 2434, Application US/09960352  
Patent No. US20020137139A1  
GENERAL INFORMATION:  
APPLICANT: Warren, Wesley C.  
APPLICANT: Tao, Nengbing  
APPLICANT: Byatt, John C.  
APPLICANT: Mathalagan, Nagappan  
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
FILE REFERENCE: 16511.006/37-21(10298)C  
CURRENT APPLICATION NUMBER: US/09/960,352  
CURRENT FILING DATE: 2001-09-24  
NUMBER OF SEQ ID NOS: 15112  
SEQ ID NO 2434  
LENGTH: 348  
TYPE: DNA  
ORGANISM: Bos taurus  
OTHER INFORMATION: Clone ID: 11-LIB3058-014-Q1-K1-C3  
US-09-960-352-2434

Query Match 78.1%; Score 16.4; DB 10; Length 348;  
Best Local Similarity 94.4%; Pred. No. 40;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 TGGCAGAACTCGCTGG 21  
DB 203 TGACGAGAACTCGCTGG 220

## RESULT 8

US-09-960-352-10478  
Sequence 10478, Application US/09960352  
Patent No. US20020137139A1  
GENERAL INFORMATION:  
APPLICANT: Warren, Wesley C.  
APPLICANT: Tao, Nengbing  
APPLICANT: Byatt, John C.  
APPLICANT: Mathalagan, Nagappan  
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
FILE REFERENCE: 16511.006/37-21(10298)C  
CURRENT APPLICATION NUMBER: US/09/960,352  
CURRENT FILING DATE: 2001-09-24  
NUMBER OF SEQ ID NOS: 15112  
SEQ ID NO 10478  
LENGTH: 422  
TYPE: DNA  
ORGANISM: Bos taurus  
OTHER INFORMATION: Clone ID: 45-LIB3058-014-Q1-K1-D2  
US-09-960-352-10478

Query Match 78.1%; Score 16.4; DB 10; Length 422;  
Best Local Similarity 94.4%; Pred. No. 41;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 TGGCAGAACTCGCTGG 21  
DB 200 TGACGAGAACTCGCTGG 217

RESULT 9  
US-09-960-352-13143

Sequence 13143, Application US/09960352  
Patent No. US20020137139A1  
GENERAL INFORMATION:  
APPLICANT: Warren, Wesley C.  
APPLICANT: Tao, Nengbing  
APPLICANT: Byatt, John C.  
APPLICANT: Mathalagan, Nagappan  
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
FILE REFERENCE: 16511.006/37-21(10298)C  
CURRENT APPLICATION NUMBER: US/09/960,352  
CURRENT FILING DATE: 2001-09-24  
NUMBER OF SEQ ID NOS: 15112  
SEQ ID NO 13143  
LENGTH: 385  
TYPE: DNA  
ORGANISM: Bos taurus  
OTHER INFORMATION: Clone ID: 56-LIB3058-047-Q1-K1-F12  
US-09-960-352-13143

Query Match 77.1%; Score 16.2; DB 10; Length 385;  
Best Local Similarity 85.7%; Pred. No. 51;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCGTGGCAGAACTCGCTGG 21  
DB 299 GCGTGGCAGAACTCGCTGG 319

## RESULT 10

US-09-963-965-4246  
Sequence 4246, Application US/09983965  
Patent No. US20020137160A1  
GENERAL INFORMATION:  
APPLICANT: Warren, Wesley C.  
APPLICANT: Tao, Nengbing  
APPLICANT: Byatt, John C.  
APPLICANT: Mathalagan, Nagappan  
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
FILE REFERENCE: 37-21(10297)C  
CURRENT APPLICATION NUMBER: US/09/983,965  
CURRENT FILING DATE: 2001-10-26  
PRIOR APPLICATION NUMBER: US 09/465,231  
PRIOR FILING DATE: 1999-12-15  
PRIOR APPLICATION NUMBER: US 60/113,678  
PRIOR FILING DATE: 1998-12-17  
NUMBER OF SEQ ID NOS: 5912  
SEQ ID NO 4246  
LENGTH: 398  
TYPE: DNA  
ORGANISM: Bos taurus  
FEATURE:  
OTHER INFORMATION: Clone ID: 61-LIB3058-040-Q1-K1-H2  
US-09-963-965-4246

Query Match 77.1%; Score 16.2; DB 10; Length 398;  
Best Local Similarity 85.7%; Pred. No. 51;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCGTGGCAGAACTCGCTGG 21  
DB 363 GCGTGGCAGAACTCGCTGG 383

## RESULT 11

US-09-960-352-4253  
Sequence 4253, Application US/09960352  
Patent No. US20020137139A1

```
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Mengding
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 4253
; LENGTH: 452
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 18-LIB34-086-Q1-E1-E5
US-09-960-352-4253

Query Match
Best Local Similarity 77.1%; Score 16.2; DB 10; Length 452;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCGTGGCAAGACTGCTGG 21
Db 17 GCGTGGCAAGAGCGCTCTGG 37

RESULT 12
US-09-764-864-769/c
; Sequence 769, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 769
; LENGTH: 454
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-864-769

Query Match
Best Local Similarity 77.1%; Score 16.2; DB 10; Length 454;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCGTGGCAAGACTGCTGG 21
Db 423 GCGTGGCAAGACTGCTGG 403

RESULT 13
US-09-960-352-5830
; Sequence 5830, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Mengding
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 5830
; LENGTH: 454
; TYPE: DNA
```

```
; ORGANISM: Bos taurus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (181),(244),(385)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: 25-LIB3058-057-Q1-K1-G1
US-09-960-352-5830

Query Match
Best Local Similarity 77.1%; Score 16.2; DB 10; Length 454;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCGTGGCAAGACTGCTGG 21
Db 96 GCGTGGCAAGAGCGCTCTGG 116

RESULT 14
US-09-764-864-354/c
; Sequence 354, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 354
; LENGTH: 498
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-864-354

Query Match
Best Local Similarity 77.1%; Score 16.2; DB 10; Length 498;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCGTGGCAAGACTGCTGG 21
Db 423 GCGTGGCAAGACTGCTGG 403

RESULT 15
US-09-962-832-224
; Sequence 224, Application US/09962832
; Patent No. US20020110821A1
; GENERAL INFORMATION:
; APPLICANT: Ebner, Reinhard
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Sign
; FILE REFERENCE: 689290-74
; CURRENT APPLICATION NUMBER: US/09/962,832
; CURRENT FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US/60/235,077
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,280
; PRIOR FILING DATE: 2000-09-25
; NUMBER OF SEQ ID NOS: 259
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 224
; LENGTH: 1443
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-962-832-224

Query Match
Best Local Similarity 77.1%; Score 16.2; DB 10; Length 1443;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCGTGGCAAGACTGCTGG 21
```

Tue Mar 25 09:37:26 2003

us-09-873-546-6.rnpb

Page 5

Db 425 |||||  
GCCTGGGCAAGAGCCCTGG 445

Search completed: March 23, 2003, 17:17:27  
Job time : 22.5425 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 23, 2003, 14:20:51 ; Search time 136.531 Seconds  
(without alignments)  
2491.051 Million cell updates/sec

File: US-09-873-546-6  
Perfect score: 21  
Sequence: 1 gcgtggcgaagactcgcctg 21

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues  
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estda:\*  
2: em\_esthum:\*  
3: em\_estln:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pla:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rtd:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19.4	92.4	448	10	AM656952 109489 MA
2	19.4	92.4	454	9	AA723009 zh0e05.s
3	19.4	92.4	470	9	AI497811 tm89f05.x
4	19.4	92.4	662	10	BB648303 BB648303
5	19.4	92.4	664	12	BG699685 BG699685
6	19.4	92.4	838	14	BO887987 BO887987

7	19	90.5	659	12	BE696122	BE696122 MRL-CT025
8	17.8	84.8	349	10	BB871148	BB871148 BB871148
9	17.8	84.8	368	10	AM437426	AM437426 76464 MAR
10	17.8	84.8	378	12	BF871691	BF871691 CM4-ET009
11	17.8	84.8	451	13	BI614393	BI614393 RH43755.5
12	17.8	84.8	494	13	BI362167	BI362167 RH46661.5
13	17.8	84.8	588	10	AV692927	AV692927 AV692927
14	17.8	84.8	590	10	AV689088	AV689088 AV689088
15	17.8	84.8	597	10	AV689087	AV689087 AV689087
16	17.8	84.8	609	13	BI607592	BI607592 RH74685.5
17	17.8	84.8	710	13	BI331932	BI331932 602984323
18	17.8	84.8	721	10	AV698268	AV698268 AV698268
19	17.8	84.8	729	13	BI247531	BI247531 602960169
20	17.8	84.8	770	12	BF257583	BF257583 BSMF001
21	17.8	84.8	793	13	BI144949	BI144949 602909601
22	17.8	84.8	796	12	BF364010	BF364010 602045538
23	17.8	84.8	798	13	BM048293	BM048293 603625501
24	17.8	84.8	802	12	BG775066	BG775066 602650030
25	17.8	84.8	862	14	BQ951805	BQ951805 AGENCOURT
26	17.8	84.8	982	12	BG743095	BG743095 602634230
27	17.8	84.8	1087	13	BM555453	BM555453 AGENCOURT
28	17.8	84.8	3079	11	AK004933	AK004933 Mus muscu
29	17.4	82.9	360	9	AJ473491	AJ473491 AJ473491
30	17.4	82.9	568	13	BJ013208	BJ013208 BJ013208
31	17.4	82.9	605	9	AA801711	AA801711 GM03370.5
32	17.4	82.9	605	9	AI063314	AI063314 GH03053.5
33	17.4	82.9	632	9	AI402550	AI402550 GH21792.5
34	17.4	82.9	654	13	BJ493532	BJ493532 BJ493532
35	17.4	82.9	678	13	BI240737	BI240737 RE37651.5
36	17.4	82.9	1144	17	CNS06R55	AL411843 t7 end of
37	17.4	81.0	449	17	AO846554	AO846554 LMAJFV1.1
38	16.8	80.0	139	13	BM293161	BM293161 EST575703
39	16.8	80.0	253	12	BF524549	BF524549 UT-R-AEO-
40	16.8	80.0	277	9	AA023366	AA023366 mh71905.r
41	16.8	80.0	300	10	BE654600	BE654600 UT-M-AJ1-
42	16.8	80.0	349	6	BM375926	BM375926 EBem06-SQ
43	16.8	80.0	353	17	AQ073553	AQ073553 EP(2)2397
44	16.8	80.0	360	14	R33532	R33532 YH79C04.r1
45	16.8	80.0	365	10	BE127748	BE127748 DEPA1497

## ALIGNMENTS

RESULT 1  
LOCUS AM656952 448 bp mRNA linear EST 25-APR-2001  
DEFINITION 109489 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.  
ACCESSION AM656952  
VERSION AM656952.1 GI:7422778  
KEYWORDS EST.  
SOURCE cow.  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
REFERENCE Smith,T.P.L., Grose,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,  
Casas,E., Wray,J.E., White,J., Cho,J., Fahnensterg,S.C., Bennett,  
G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,  
Peters,G., Holt,I., Karaycheva,S., Liang,F., Quackenbush,J. and  
Keefe,J.W.  
Sequence evaluation of four pooled-tissue normalized bovine cDNA  
libraries and construction of a gene index for cattle  
Genome Res. 11 (4), 626-630 (2001)  
CONTACT: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smith@email.marc.usda.gov  
Single pass sequencing. Bases called and alt-trimmed with phred







Query Match 92.4%; Score 19.4; DB 12; Length 664;  
 Best Local Similarity 95.2%; Pred. No. 1.9e+02;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCGTGGCAAGAACTCGCTGG 21  
 |||||||||  
 Db 187 GCGTGGCAAGAACTCGCTGG 207

RESULT 6  
 LOCUS BQ887987 838 bp mRNA linear EST 16-AUG-2002  
 DEFINITION AGENCOURT.8742650 NIH\_MGC\_129 Mus musculus cDNA clone IMAGE:6390947  
 5' mRNA sequence.  
 ACCESSION BQ887987  
 VERSION BQ887987.1 GI:22280001  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 838)  
 NIH-MGC http://mgi.nhl.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cga@bbs-remail.nih.gov  
 Tissue Procurement: Susan L. Sullivan, PhD.  
 cDNA Library Preparation: Resgen, Invitrogen Corp.  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLM13878 row: 9 column: 12  
 High quality sequence stop: 398.  
 Location/Qualifiers  
 1..838  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone\_image="6390947"  
 /clone\_lib="NIH\_MGC\_129"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: olfactory epithelium; Vector:  
 pCMV-Sport6.1 ccdB; Site\_1: EcoRV; Site\_2: NotI. Cloned  
 unidirectionally. Primer: Oligo dt. Average insert size  
 2.2 kb. Constructed by Resgen, Invitrogen Corp. Note: this  
 is a NIH\_MGC Library."  
 BASE COUNT 150 a 277 c 227 g 184 t  
 ORIGIN

Query Match 92.4%; Score 19.4; DB 14; Length 838;  
 Best Local Similarity 95.2%; Pred. No. 2.1e+02;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCGTGGCAAGAACTCGCTGG 21  
 |||||||||  
 Db 58 GCGTGGCAAGAACTCGCTGG 78

RESULT 7  
 LOCUS BE696122 659 bp mRNA linear EST 11-SEP-2000  
 DEFINITION MRI-C70258-040700-002-F04 C70258 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION BE696122  
 VERSION BE696122.1 GI:10083282  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 659)

AUTHORS  
 Dias Neto, E., Garcia Correia, R., Verjovski-Almeida, S., Briones, M. R.,  
 Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F.,  
 Goldman, G. H., Carvalho, A. F., Matsukuma, A., Bala, G. S., Simpson, D. H.,  
 Brunstein, A., de Oliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare,  
 M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and  
 Simpson, A. J. G.

TITLE  
 Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags

JOURNAL  
 MEDLINE  
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 20202663

COMMENT  
 Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?l=ct2-MRI-C70258-040  
 700-002-f04&ct3=2000-07-04&ct4=1)  
 Seq primer: puc 18 forward  
 High quality sequence start: 10  
 High quality sequence stop: 599.  
 Location/Qualifiers  
 1..659  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="C70258"  
 /dev\_stage="Adult"  
 /note="Organ: colon; Vector: puc18; Site\_1: SmaI; Site\_2:  
 SmaI; A mini-library was made by cloning products derived  
 from ORESTES PCR (U.S. Letters Patent application No. 196  
 ,716 - Ludwig Institute for Cancer Research) profiles  
 into the puc 18 vector. Reverse transcription of tissue  
 mRNA and cDNA amplification were performed under low  
 stringency conditions."

BASE COUNT 166 a 157 c 174 g 162 t  
 ORIGIN

Query Match 90.5%; Score 19; DB 12; Length 659;  
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GTGGGCAAGAACTCGCTGG 21  
 |||||||||  
 Db 37 GTGGGCAAGAACTCGCTGG 55

RESULT 8  
 LOCUS BB871148 349 bp mRNA linear EST 27-NOV-2001  
 DEFINITION BB871148 RIKEN full-length enriched, 1 month neonate cerebellum Mus  
 musculus cDNA clone G630027D18 5', mRNA sequence.  
 ACCESSION BB871148  
 VERSION BB871148.1 GI:17117358  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 349)  
 Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T.,  
 Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Imotani, K., Ishii,  
 Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Matsuyama, T.,  
 Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T.,  
 Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K.,  
 Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa,  
 A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T.,  
 Wataniki, A., Yasunishi, A., Yumatsugu, M. and Hayashizaki, Y.,  
 RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al.  
 2001)

JOURNAL  
COMMENT

Unpublished (2001)  
 Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic  
 Sciences Center(GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@gs.riken.go.jp,  
 URL: http://genome.gsc.riken.go.jp/  
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh  
 M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new  
 genes. Genome Res. 10 (10), 1617-1630 (2000)  
 Nagl, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,  
 Watanabe, M., Itoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura  
 S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and  
 Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
 10 (11), 1757-1771 (2000)  
 Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara  
 Y. and Hayashizaki, Y.  
 Computer-based methods for the mouse full-length cDNA  
 encyclopedia: real-time sequence clustering for construction of a  
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 Please visit our web site (http://genome.gsc.riken.go.jp) for  
 further details.  
 e mouse tissues.

## FEATURES

source

Location/Qualifiers

1. 349  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone\_id="G63002/D18"  
 /clone\_lib="RIKEN full-length enriched, 1 month neonate  
 cerebellum"  
 /tissue\_type="cerebellum"  
 /dev\_stage="1 month neonate"  
 /note="pooled tissues; (tissue\_type=cerebellum,  
 dev\_stage=16 days neonate, sex=mixed),  
 (tissue\_type=cerebellum, dev\_stage=0 day neonate,  
 sex=mixed), (tissue\_type=hippocampus, dev\_stage=adult,  
 sex=male), (tissue\_type=whole body, dev\_stage=9 days  
 embryo, sex=mixed), (tissue\_type=lung, dev\_stage=13 days  
 embryo, sex=mixed)"  
 BASE COUNT 68 a 110 c 114 g 57 t  
 RIGIN

Query Match 84.8%; Score 17.8; DB 10; Length 349;  
 Best Local Similarity 90.5%; Pred. No. 8.1e+02;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GCGTGGCAGAGACTCGCTGG 21  
 Db 149 GCGTGGCAGAGACTCGCTGG 169

RESULT 9 368 bp mRNA linear EST 25-APR-2001  
 LOCUS AM437426  
 DEFINITION 78464 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.  
 ACCESSION AM437426  
 VERSION AM437426.1 GI:6972732  
 KEYWORDS EST.  
 SOURCE BOS taurus  
 ORGANISM Bos taurus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 Bovinae; Bos.

REFERENCE 1 (bases 1 to 368)  
 Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,

## TITLE

JOURNAL  
MEDLINE  
COMMENT

Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett  
 G.L., Heaton, M.P., Laegreid, W.M., Rohrer, G.A., Chitko-McKown, C.G.,  
 Pereira, G., Holt, I., Karanycheva, S., Liang, F., Quackenbush, J. and  
 Keeler, J.W.  
 Sequence evaluation of four pooled-tissue normalized bovine cDNA  
 libraries and construction of a gene index for cattle  
 Genome Res. 11 (4), 626-630 (2001)  
 21180013  
 Contact: Smith TPL  
 USDA, ARS, US Meat Animal Research Center  
 PO Box 166, Clay Center, NE 68933-0166, USA  
 Tel: 402 762 4366  
 Fax: 402 762 4390  
 Email: smith@emall.marc.usda.gov  
 Single pass sequencing. Bases called and trimmed with phred  
 v0.980904.e. Vector identified by cross\_match with the -m1nscore 20  
 and -mismatch 12 options.  
 PCR Primers  
 FORWARD: AGGAACACGCTATGACCAT  
 BACKWARD: GTTTCACGTCACGACG  
 Plate: 42 row: P column: 2  
 Seq primer: ATTAGCTGACACTATAG.  
 Location/Qualifiers

## FEATURES

source

1. 368  
 /organism="Bos taurus"  
 /db\_xref="taxon:9913"  
 /clone\_lib="MARC 1BOV"  
 /tissue\_type="pooled"  
 /lab\_host="DH10B"  
 /note="Vector: pCMV SPORT6; Site\_1: NotI; Site\_2: SalI;  
 Library made from pooled tissue from lymph node, ovary,  
 fat, hypothalamus, and pituitary."  
 BASE COUNT 61 a 123 c 118 g 66 t  
 ORIGIN

Query Match 84.8%; Score 17.8; DB 10; Length 368;  
 Best Local Similarity 90.5%; Pred. No. 8.2e+02;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GCGTGGCAGAGACTCGCTGG 21  
 Db 205 GCGTGGCAGAGACTCGCTGG 225

RESULT 10 378 bp mRNA linear EST 17-JAN-2001  
 LOCUS BF871691/c  
 DEFINITION CM4-ET0096-011100-400-f10 ET0096 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION BF871691  
 VERSION BF871691.1 GI:12261821  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 378)

Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,  
 Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare  
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
 Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags  
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 20202663  
 Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil

tel: +55-11-2704922  
 fax: +55-11-2707001

Email: asimpson@ludwig.org.br  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM4&tz=CM4-ET0096-011100-400-f106t3-2000-11-01&t4=1)

Seq primer: puc 18 forward  
High quality sequence start: 122  
High quality sequence stop: 225.

## FEATURES

Location/Qualifiers

1..378

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_idb="ET0096"

/dev\_stage="Adult"

/note="Organ: lung,tumor. Vector: puc18; Site\_1: SmaI;

Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (O.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 90 a 121 c 94 g 73 t

## ORIGIN

Query Match 84.8%; Score 17.8; DB 12; Length 378;  
Best Local Similarity 90.5%; Pred. No. 8.3e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GCCTGGCAGAGCTCGCTG 21

Db 250 GCCTGGCAGAGCTCGCTG 230

## RESULT 11

BI614393

## LOCUS

BI614393 451 bp mRNA linear EST 07-SEP-2001  
RH43755.5prime RH Drosophila melanogaster normalized Head pf1c-1  
Drosophila melanogaster cDNA clone RH43755 5 similar to CG8500:  
Fban0008500 GO:[] located on: 3R 85E10-85E10; 08/18/2001, mRNA  
sequence.

## ACCESSION

BI614393.1 GI:15509918

## VERSION

## KEYWORDS

## SOURCE

ORGANISM

EST.

fruit fly.

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 451)

Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B., Carlson  
,J., Chamez,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George  
,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Mista,S.,  
Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,  
Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Celinker,S. and Rubin  
,G.M.

## REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu

hit genomic AE003684: arm:3R [5513047,5732625]

estimated-cyto:85E6-85E2: 08/18/2001

Plate: RH 437 row: E column: 7

High quality sequence stop: 450.

Location/Qualifiers

1..451

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone\_idb="RH43755"

/clone\_idb="RH Drosophila melanogaster normalized Head

pf1c-1"  
/sex="male and female"  
/dev\_stage="Adult"  
/lab\_host="DH5-alpha Tona"  
/note="Organ: head; Vector: pf1c1; Site\_1: XhoI; Site\_2:  
BamHI; Library was kindly generated by Piero Carninci at  
the RIKEN. The library was normalized and excised using  
Cre recombinase. Plasmid cDNA library."

## BASE COUNT

122 a 132 c 115 g 82 t

## ORIGIN

Query Match 84.8%; Score 17.8; DB 13; Length 451;  
Best Local Similarity 90.5%; Pred. No. 8.8e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GCCTGGCAGAGCTCGCTG 21

Db 270 GCCTGGCAGAGCTCGCTG 290

## RESULT 12

BI362167

## LOCUS

BI362167 494 bp mRNA linear EST 01-AUG-2001  
RE46651.5prime RE Drosophila melanogaster normalized Embryo pf1c-1  
Drosophila melanogaster cDNA clone RE46651 5 similar to CG8500:  
Fban0008500 located on: 3R 85E10-85E10; 05/13/2001, mRNA sequence.

## ACCESSION

BI362167

## VERSION

## KEYWORDS

## SOURCE

ORGANISM

EST.

fruit fly.

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 494)

Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B., Carlson  
,J., Chamez,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George  
,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Mista,S.,  
Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,  
Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Celinker,S. and Rubin  
,G.M.

## REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu

hit genomic AE003684: arm:3R [5513047,5732625]

estimated-cyto:85E6-85E2: 05/13/2001

Plate: RE 466 row: F column: 1

High quality sequence stop: 429.

Location/Qualifiers

1..494

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone\_idb="RE46651"

/clone\_idb="RE Drosophila melanogaster normalized Embryo

pf1c-1"

/sex="male and female"

/dev\_stage="0-24 hours mixed stage embryonic"

/lab\_host="DH5-alpha Tona"

/note="Organ: embryo; Vector: pf1c1; Site\_1: XhoI; Site\_2:

BamHI; Library was kindly generated by Piero Carninci at

the RIKEN. The library was normalized and excised using

Cre recombinase. Plasmid cDNA library."

BASE COUNT 133 a 148 c 124 g 89 t

ORIGIN

Query Match 84.8%; Score 17.8; DB 13; Length 494;  
Best Local Similarity 90.5%; Pred. No. 9.1e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GCCTGGCAGAACTCGCTGG 21  
|||||  
Db 313 GCCTGGCAGAACTCGCTGG 333

RESULT 13  
AV692927 588 bp mRNA linear EST 16-JAN-2002  
LOCUS AV692927 GKC Homo sapiens cDNA clone GKCD03 5', mRNA sequence.  
ACCESSION AV692927  
VERSION AV692927.1 GI:10294790  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 588)  
XU,X., HUANG,J., XU,Z., QIAN,B., ZHU,Z., YAN,Q., CAI,T., ZHANG,X.,  
XIAO,H., QU,J., LIU,F., HUANG,Q., CHENG,Z., LI,N., DU,J., HU,W.,  
SHEN,K., LU,G., FU,G., ZHONG,M., XU,S., GU,W., HUANG,W., ZHAO,X.,  
HU,G., GU,J., CHEN,Z., and HAN,Z.  
Insight into hepatocellular carcinogenesis at transcriptome level  
by comparing gene expression profiles of hepatocellular carcinoma  
with those of corresponding noncancerous liver  
Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)  
21625106

JOURNAL  
MEDLINE  
COMMENT  
Contact: Zeguang Han  
Chinese National Human Genome Center at Shanghai  
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai  
201203, P. R. China  
Tel: 86-21-50801919(ex.45)  
Fax: 86-21-50801922  
Email: hanzg@chgc.sh.cn  
This clone is available at CHGC in Shanghai.

FEATURES  
source  
location/Qualifiers  
1..588  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_1lb="GKCD03"  
/clone\_1lb="GKC"  
/tissue\_type="hepatocellular carcinoma"  
/dev\_stage="Adult"  
/lab\_host="SOLR"  
/note="Vector: pBluescript sk(-); Site\_1: EcoRI; Site\_2:  
XhoI"

BASE COUNT 95 a 198 c 185 g 110 t

ORIGIN

Query Match 84.8%; Score 17.8; DB 10; Length 588;  
Best Local Similarity 90.5%; Pred. No. 9.6e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GCCTGGCAGAACTCGCTGG 21  
|||||  
Db 82 GCCTGGCAGAACTCGCTGG 102

RESULT 14  
AV689088 590 bp mRNA linear EST 16-JAN-2002  
LOCUS AV689088 GKC Homo sapiens cDNA clone GKCD2B06 5', mRNA sequence.  
ACCESSION AV689088  
VERSION AV689088.1 GI:10290951  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 590)  
XU,X., HUANG,J., XU,Z., QIAN,B., ZHU,Z., YAN,Q., CAI,T., ZHANG,X.,  
XIAO,H., QU,J., LIU,F., HUANG,Q., CHENG,Z., LI,N., DU,J., HU,W.,  
SHEN,K., LU,G., FU,G., ZHONG,M., XU,S., GU,W., HUANG,W., ZHAO,X.,

TITLE  
Hu,G., Gu,J., Chen,Z. and Han,Z.  
Insight into hepatocellular carcinogenesis at transcriptome level  
by comparing gene expression profiles of hepatocellular carcinoma  
with those of corresponding noncancerous liver  
Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)  
21625106

JOURNAL  
MEDLINE  
COMMENT  
Contact: Zeguang Han  
Chinese National Human Genome Center at Shanghai  
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai  
201203, P. R. China  
Tel: 86-21-50801919(ex.45)  
Fax: 86-21-50801922  
Email: hanzg@chgc.sh.cn  
This clone is available at CHGC in Shanghai.

FEATURES  
source  
location/Qualifiers  
1..590  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_1lb="GKCD2B06"  
/clone\_1lb="GKC"  
/tissue\_type="hepatocellular carcinoma"  
/dev\_stage="Adult"  
/lab\_host="SOLR"  
/note="Vector: pBluescript sk(-); Site\_1: EcoRI; Site\_2:  
XhoI"

BASE COUNT 92 a 198 c 183 g 112 t 5 others

ORIGIN

Query Match 84.8%; Score 17.8; DB 10; Length 590;  
Best Local Similarity 90.5%; Pred. No. 9.6e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GCCTGGCAGAACTCGCTGG 21  
|||||  
Db 82 GCCTGGCAGAACTCGCTGG 102

RESULT 15  
AV689087 597 bp mRNA linear EST 16-JAN-2002  
LOCUS AV689087 GKC Homo sapiens cDNA clone GKCD2B05 5', mRNA sequence.  
ACCESSION AV689087  
VERSION AV689087.1 GI:10290950  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 597)  
XU,X., HUANG,J., XU,Z., QIAN,B., ZHU,Z., YAN,Q., CAI,T., ZHANG,X.,  
XIAO,H., QU,J., LIU,F., HUANG,Q., CHENG,Z., LI,N., DU,J., HU,W.,  
SHEN,K., LU,G., FU,G., ZHONG,M., XU,S., GU,W., HUANG,W., ZHAO,X.,  
HU,G., GU,J., CHEN,Z., and HAN,Z.  
Insight into hepatocellular carcinogenesis at transcriptome level  
by comparing gene expression profiles of hepatocellular carcinoma  
with those of corresponding noncancerous liver  
Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)  
21625106

JOURNAL  
MEDLINE  
COMMENT  
Contact: Zeguang Han  
Chinese National Human Genome Center at Shanghai  
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai  
201203, P. R. China  
Tel: 86-21-50801919(ex.45)  
Fax: 86-21-50801922  
Email: hanzg@chgc.sh.cn  
This clone is available at CHGC in Shanghai.

FEATURES  
source  
location/Qualifiers  
1..597  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_1lb="GKCD2B05"  
/clone\_1lb="GKC"  
/tissue\_type="hepatocellular carcinoma"

/dev\_stage="Adult"  
/lab\_host="SOLR"  
/note="Vector: pBluescript sk(-); Site\_1: EcoRI; Site\_2:  
XhoI"  
BASE COUNT 96 a 196 c 182 g 117 t 6 others  
ORIGIN

Query Match 84.8%; Score 17.8; DB 10; Length 597;  
Best Local Similarity 90.5%; Pred. No. 9.6e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 GCCTGGGCAAGACTGCTGG 21  
|||  
DB 81 GCCTGGGCAAGACTGCTGG 101

Search completed: March 23, 2003, 17:12:35  
Job time : 140.531 secs

GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 23, 2003, 13:56:37 ; Search time 694.166 Seconds  
(without alignments)  
1936.775 Million cell updates/sec

Title: US-09-873-546-4

Perfect score: 597

Sequence: 1 atgccgcagacagatgaacga.....gcaatgcacccctcatgtga 597

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Optical number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.Geneseq\_101002:.\*  
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	583.4	97.7	1249	24	ABO72525 Human MDDT encodin
2	582.8	97.6	3061	24	ABO72648 Human MDDT encodin
3	380.8	63.8	1087	22	AA160838 Human polynucleoti
4	380.8	63.8	1108	22	AA159052 Human polynucleoti
5	380.8	63.8	4167	24	AD37605 Human intracellula
6	303.2	50.8	702	23	AB114159 Drosophila melanog
7	222	37.2	1286	22	AA583350 Human GTP-binding
8	222	37.2	1561	22	AA15447 Human cDNA sequenc
9	222	37.2	3403	23	AA587856 DNA encoding novel

10	220.4	36.9	7397	19	AAV60578 Human tumour suppl
11	200.8	33.6	1481	19	AAV60577 Human tumour suppl
12	175.4	29.4	1144	22	AAH90075 Human bone marrow
13	175.4	29.4	4933	22	AAK81235 Human immune/thema
14	175.4	29.4	4934	22	AAK81234 Human immune/thema
15	166.4	27.9	558	11	AAQ03212 RAP2 Gene encoding
16	166.4	27.9	3300	22	AA158569 Human polynucleoti
17	160.6	26.9	2951	23	AB114158 Drosophila melanog
18	153.2	25.7	1170	23	AB103763 Drosophila melanog
19	153.2	25.7	3554	23	AB103762 Drosophila melanog
20	152.6	25.6	837	21	AAZ36913 DNA encoding a hom
21	152.6	25.6	2832	24	AB192076 Human Tumour Endot
22	152.6	25.6	2973	24	AB192087 Human Tumour Endot
23	152.6	25.6	3427	24	ABK71563 Human dltip polynu
24	144.8	24.3	3020	24	AB192134 Mouse Tumour Endot
25	139.4	23.4	2223	22	AAH41597 Human H-Ras, c-Raf
26	139.4	23.4	2223	24	AB150184 Human H-Ras, c-Raf
27	139.4	23.4	2223	24	AB150216 Human H-Ras, c-Raf
28	139.4	23.4	2295	24	AB150201 Human H-Ras, c-Raf
29	139.4	23.4	2295	24	AB150233 Human H-Ras, c-Raf
30	137.8	23.1	570	21	AAZ50192 Ras gene for TDNE
31	137.8	23.1	570	21	AAZ50192 Human Ras cDNA, H
32	137.8	23.1	2223	22	AAH41609 Human H-Ras, c-Raf
33	137.8	23.1	2223	24	AB150196 Human H-Ras, c-Raf
34	137.8	23.1	2223	24	AB150228 Human H-Ras, c-Raf
35	137.4	23.0	657	24	ABK14232 Human related RAS
36	137.4	23.0	2223	22	AAH41606 Human R-Ras, c-Raf
37	137.4	23.0	2223	24	AB150193 Human R-Ras, c-Raf
38	137.4	23.0	2223	24	AB150225 Human R-Ras, c-Raf
39	137.2	22.6	1933	23	AB151513 Drosophila melanog
40	134.8	22.6	846	21	AAZ36892 cDNA encoding an a
41	134.8	22.6	1740	21	AAZ36894 cDNA encoding an a
42	134.8	22.6	1801	21	AAZ36893 cDNA encoding an a
43	134.8	22.6	1841	20	AAZ23022 Human Kd312 polype
44	133.2	22.3	1776	21	AAA49177 cDNA encoding huma
45	127.4	21.3	561	24	AB158432 Rat oncogenic cyto

#### ALIGNMENTS

RESULT 1	ABO72525	ABO72525 standard; cDNA; 1249 BP.
ID	ABO72525	
XX	ABO72525	
AC	ABO72525	
XX	03-SEP-2002 (first entry)	
DT	03-SEP-2002	
XX	Human MDDT encoding cDNA SEQ ID NO 77.	
DE	Human MDDT encoding cDNA SEQ ID NO 77.	
XX	Human: MDDT; disease detection and treatment molecule polynucleotide;	
KW	proliferative disorder; hepatitis; psoriasis; cancer; AIDS;	
KW	autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;	
KW	rheumatoid arthritis; transgenic; gene therapy; antitumor; antitumor;	
KW	hepatotoxic; antineoplastic; antiproliferative; antiproliferative; antiproliferative;	
KW	antiallergic; antineoplastic; antineoplastic; antineoplastic; antineoplastic;	
KW	neuroprotective; antineoplastic; antineoplastic; antineoplastic; antineoplastic;	
XX	Homo sapiens.	
OS	Homo sapiens.	
XX	WO200240715-A2.	
PN	WO200240715-A2.	
XX	23-MAY-2002.	
PD	23-MAY-2002.	
XX	06-SEP-2001; 2001WO-0527628.	
PF	06-SEP-2001; 2001WO-0527628.	
XX	06-SEP-2000; 2000US-230505P.	
PR	06-SEP-2000; 2000US-230505P.	
PR	06-SEP-2000; 2000US-230514P.	
PR	06-SEP-2000; 2000US-230515P.	
PR	06-SEP-2000; 2000US-230517P.	
PR	06-SEP-2000; 2000US-230518P.	
PR	06-SEP-2000; 2000US-230519P.	





PR 06-SEP-2000; 2000US-230865P.  
 PR 06-SEP-2000; 2000US-230988P.  
 PR 06-SEP-2000; 2000US-230989P.  
 PR 07-SEP-2000; 2000US-230951P.  
 PR 07-SEP-2000; 2000US-231163P.  
 PR 07-SEP-2000; 2000US-231167P.  
 XX  
 PA (INCYTE GENOMICS INC.)  
 XX  
 PI Jackson S., Lincoln SE., Altus CM., Dufour GE., Chalup MS., Hillman JL.,  
 PI Jones AL., Yu JT., Wright RJ., Gietzen D., Liu TF., Yap PE., Dahl CR.,  
 PI Momiyama MG., Bradley DL., Rohatgi SD., Harris B., Roseberry AM.,  
 PI Gestlin EH., Peralta CH., David MH., Panzer SR., Flores V., Daffo A.,  
 PI Marwaha R., Chen AJ., Chang SC., Au AP., Iman RR.  
 XX  
 DR WPI: 2002-527544/56.  
 DR P-PSDB: ABP51431.  
 XX  
 PT Novel human disease detection and treatment polypeptide, useful in  
 PT diagnosis, prevention or treatment of cell proliferative disorders e.g.  
 PT arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder  
 PT e.g. AIDS  
 XX  
 PS Claim 1; Page 414; 618bp; English.  
 XX  
 CC The invention relates to an isolated human disease detection and  
 CC treatment (MDPT) polypeptide (I) selected from a polypeptide having a  
 CC sequence selected from 254 sequences (ABP51231-ABP51484) given in the  
 CC specification, a naturally occurring polypeptide comprising a sequence  
 CC having at least 90% identity to (I) or a biologically active or  
 CC immunogenic fragment of (I). (I) is useful for screening a compound for  
 CC effectiveness as an agonist or antagonist, for screening a compound that  
 CC specifically binds (I) or modulates the activity of (I), and for  
 CC preparing a polyclonal or monoclonal antibody by hybridoma technology.  
 CC Nucleic acids (II) (ABQ72449-ABQ72700) encoding (I) are useful for  
 CC screening a compound for effectiveness in altering expression of a target  
 CC polynucleotide comprising. Oligonucleotides and antibodies are useful for  
 CC detecting MDP in a sample or for assessing toxicity of a test compound,  
 CC in a diagnostic test for a condition or a disease associated with the  
 CC expression of MDP in a biological sample, for detecting (I) in a sample,  
 CC and for purifying (I) from a sample. A composition comprising (I), an  
 CC agonist or antagonist is useful for treating a disease or condition  
 CC associated with decreased or increased expression of functional MDP.  
 CC (I) or (II) are useful for diagnosing, treating or preventing disorders  
 CC associated with aberrant expression of MDP, where the disorders are  
 CC selected from a cell proliferative disorder such as arteriosclerosis,  
 CC cirrhosis, hepatitis, psoriasis, and cancer and an  
 CC autoimmune/inflammatory disorder such as AIDS, Addison's disease,  
 CC allergy, anaemia, asthma, atherosclerosis, gout, multiple sclerosis or  
 CC rheumatoid arthritis. (II) are useful for creating knockin humanised  
 CC animals or transgenic animals to model human diseases, in somatic or  
 CC germine gene therapy, to generate a transcript image of a tissue or cell  
 CC type, for detecting differences in the chromosomal location due to  
 CC translocation or inversion among normal, carrier or affected individuals  
 CC and as hybridisation probes for mapping naturally occurring genomic  
 CC sequences.  
 XX  
 XX Sequence 3061 BP; 606 A; 840 C; 927 G; 688 T; 0 other;  
 XX  
 S0  
 Query Match 97.6%; Score 582.8; DB 24; Length 3061;  
 Best Local Similarity 99.5%; Pred. No. 2.4e-103;  
 Matches 595; Conservative 0; Mismatches 2; Indels 1; Gaps 1;  
 OY 1 ATGCCGGAACAGAGTACGATTACCGCGTGGTTCGGGGGGGGGGGGTGGGCAAG 60  
 DB 51 ATGCCGGAACAGAGTACGATTACCGCGTGGTTCGGGGGGGGGGGGTGGGCAAG 110  
 OY 61 AGCTGCTGCTGCTGCGCTTGTGTAAGGACAGCTTCCGACACCTACATCCCAACATC 120  
 DB 111 AGCTGCTGCTGCTGCGCTTGTGTAAGGACAGCTTCCGACACCTACATCCCAACATC 170  
 OY 121 GAGGACACTTACCGGACAGGTATCAGCTGCGACAAAGACGCTGACGCTGACATCACA 180  
 DB 121 GAGGACACTTACCGGACAGGTATCAGCTGCGACAAAGACGCTGACGCTGACATCACA 180

DB 171 GAGGACACTTACCGGACAGGTATCAGCTGCGACAAAGACGCTGTCGACCGCTGACATCACA 230  
 OY 181 GAGGACACTTACCGGACAGGTATCAGCTGCGACAAAGACGCTGTCGACCGCTGACATCACA 240  
 DB 231 GAGGACACTTACCGGACAGGTATCAGCTGCGACAAAGACGCTGTCGACCGCTGACATCACA 290  
 OY 241 GAGGACACTTACCGGACAGGTATCAGCTGCGACAAAGACGCTGTCGACCGCTGACATCACA 300  
 DB 291 GAGGACACTTACCGGACAGGTATCAGCTGCGACAAAGACGCTGTCGACCGCTGACATCACA 350  
 OY 301 TACAAAGCTCATGCTGACATCAGTCAAGGACAGCTGAGAGACATCCCGTATGCTGAGGAC 360  
 DB 351 TACAAAGCTCATGCTGACATCAGTCAAGGACAGCTGAGAGACATCCCGTATGCTGAGGAC 410  
 OY 361 AACAAAGCTCATGCTGACATCAGTCAAGGACAGCTGAGAGACATCCCGTATGCTGAGGAC 420  
 DB 411 AACAAAGCTCATGCTGACATCAGTCAAGGACAGCTGAGAGACATCCCGTATGCTGAGGAC 470  
 OY 421 GAGTGAAGTGGCGCTTATGAGACCTGCGCAAGATGAACTACAGCTCAAGAGAG-CT 479  
 DB 471 GAGTGAAGTGGCGCTTATGAGACCTGCGCAAGATGAACTACAGCTCAAGAGAGTCT 530  
 OY 480 CTTCAGAGAGCTGCTGACGCTGAGAGAGCGCGCGAACAATGAGCTCAACATGACAGCGCAA 539  
 DB 531 CTTCAGAGAGCTGCTGACGCTGAGAGAGCGCGCGAACAATGAGCTCAACATGACAGCGCAA 590  
 OY 540 GCGCTCCGGGAAGCAG 597  
 DB 591 GCGCTCCGGGAAGCAG 648  
 XX  
 XX RESULT 3  
 XX AA160838  
 XX ID AA160838 standard; cDNA; 1087 BP.  
 XX  
 XX AA160838;  
 XX  
 XX 22-OCT-2001 (first entry)  
 XX  
 XX Human polynucleotide seq ID NO 4827.  
 XX  
 XX Human; noctropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 XX peripheral nervous system; neuropathy; central nervous system; CNS;  
 XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 XX chemokine; thrombolytic; drug screening; arthritis; inflammation;  
 XX leukaemia; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200153312-A1.  
 XX  
 PD 26-JUL-2001.  
 XX  
 XX 26-DEC-2000; 2000WO-US34263.  
 XX  
 XX 21-JAN-2000; 2000US-0488725.  
 XX 25-APR-2000; 2000US-0552317.  
 XX 09-JUL-2000; 2000US-0598042.  
 XX 19-JUL-2000; 2000US-0620312.  
 XX 03-AUG-2000; 2000US-0653450.  
 XX 14-SEP-2000; 2000US-0662191.  
 XX 19-OCT-2000; 2000US-0693036.  
 XX 29-NOV-2000; 2000US-0727344.  
 XX  
 XX (HYSE-) HYSEQ INC.  
 XX  
 XX Tang YF., Liu C., Asundi V., Chen R., Ma Y., Qian XB., Ren F., Wang D.;  
 XX Wang J., Wang Z., Wehrman T., Xu C., Xue AJ., Yang Y., Zhang J.;  
 XX Zhao QH., Zhou P., Goodrich R., Drmanac RT;  
 XX  
 XX WPI: 2001-442253/47.  
 XX P-PSDB: AAM41682.





OY 418 CAAGAGTGAAGTGGCTTTTCATGAGACCTGGCCCAAGATGACTACAGCTCAAGGAG 477  
 DB 581 CCACATGGAAGTGTCTTTCATGAGACCTGACGCAAGCTCAACCATTAAGCAAGGAG 640  
 OY 478 CTCTTCAGAGAGCTGTGAGCGGCGCCGGAACATGAGCCCTCAACATCGAGCGC 537  
 DB 641 CTTTTCAGAGAGCTGTGAGCGGCGCCGGAACATGAGCCCTCAACATCGAGCGG 700  
 OY 538 AAGCGCTCGGGAAG 597  
 DB 701 AAAAAG 760  
 RESULT 6  
 ABL14159  
 ID ABL14159 standard; cDNA; 702 BP.  
 AC ABL14159;  
 XX  
 XX 26-MAR-2002 (first entry)  
 KM Drosophila melanogaster expressed polynucleotide seq ID NO 36959.  
 KM Drosophila: developmental biology; cell signalling; insecticide;  
 KM pharmaceutical; gene; ss.  
 XX Drosophila melanogaster.  
 OS WO200171042-A2.  
 PN 27-SEP-2001.  
 PD 23-MAR-2001; 2001WO-US09231.  
 PE 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 XX (PEKE) PE CORP NY.  
 PA Venter JC, Adams M, Li PMD, Myers EW;  
 PI WPI; 2001-656860/75.  
 DR P-PSDB; ABB70056.  
 XX  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 XX Claim 1; SEQ ID NO 36959; 21pp + Sequence Listing; English.  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX Sequence 702 BP; 185 A; 187 C; 220 G; 110 T; 0 other;  
 SQ  
 Query Match 50.8%; Score 303.2; DB 23; Length 702;  
 Best Local Similarity 72.4%; Pred. No. 1.3e-49;  
 Matches 424; Conservative 0; Mismatches 153; Indels 9; Gaps 2;  
 OY 4 CCGGAGAGAGATGAGTATACCGCGTGTGTTGGGGCGGCGGTGGGCAAGAC 63  
 DB 37 CCGGAGAGAGATGAGTATACCGCGTGTGTTGGGGCGGCGGTGGGCAAGAC 96  
 OY 64 TCGGTGTGCTCGCTTGTGAAGGACGTTCCGGAGACCTACATCCCAACATCGAG 123

DB 97 TCACCTGCTGTGGCTTCAATAAAGGACGCTTCGGGAAAGCTACATCCCAACATCGAG 156  
 OY 124 GACACCTACCGGAGGATGATCAGCTGCAGCAAGAGGCTGTGGCGGAGATCAGAC 183  
 DB 157 GATACGTACAGACAGGATCAATAGCTGCACAGAGACATCTGCGCTGCAATACGAGAC 216  
 OY 184 ACCACCGGAGCCAGCAGATTTCCCGGCGATGACAGCGCTGTCTCATTCAGAGGCGCAGCC 243  
 DB 217 ACCACGGAGATCGCATCAATTTCCCGGCGATGACAGCGCTGTCTCATTCAGAGGCGCAGCC 276  
 OY 244 TTCTATCTGTGTGTCTTCCGTCACACAGCAAGTCCCTGGAGAGCTGGGGCCCATCTAC 303  
 DB 277 TTCTATCTGTGTGTGTCTTCCGTCACACAGCAAGTCCCTGGAGAGCTGGGGCCCATCTAC 336  
 OY 304 AACGTCATGTCAGATCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 360  
 DB 337 GCGCTTATCAAGAGAGCTGAAGAGGCGCGGAGATTTCCCAACATACCTGAATGTTGGTGGG 396  
 OY 361 AACAGTGGATGAGAC-----GCAGCGGAGGTGAGACAGCGCGGAGGCGGAGGCGGTG 414  
 DB 397 AACAGTGGATGAGACCGCGGAGGTGAGAGAGTCTCCCAAGCGGAGGAGGAGGAGGAGG 456  
 OY 415 GCCCAAGATGAGAGTGGCTTTTCATGAGAGAGCTGGGCCAAGTGAACATCAACATCGAG 474  
 DB 457 GCGACCACTCGAGCATATCTGTCATGAGAGCTGGGCCAAGCATCAATGTGACC 516  
 OY 475 GAGCTTTCAGAGAGCTGCTGAG 534  
 DB 517 GAGCTTTCAGAGAGCTGCTGAG 576  
 OY 535 GCGACGCTCGGGAAG 580  
 DB 577 ACCAAG 622  
 RESULT 7  
 AAF58350  
 ID AAF58350 standard; cDNA; 1286 BP.  
 AC AAF58350;  
 XX  
 XX 19-APR-2001 (first entry)  
 DT  
 XX Human GTP-binding associated protein #50 coding sequence.  
 DE  
 XX Human; guanosine triphosphate binding associated protein; GTP; GBAF;  
 KM inflammation; AIDS; Addison's disease; anaemia; arteriosclerosis; asthma;  
 KM autoimmune disorder; hepatitis; multiple sclerosis; cancer; diabetes;  
 KM osteoporosis; psoriasis; ss.  
 XX  
 XX Homo sapiens.  
 OS  
 XX WO200105970-A2.  
 PN 25-JAN-2001.  
 PD 19-JUL-2000; 2000WO-US19698.  
 PE 19-JUL-1999; 99US-0144595.  
 PR 23-AUG-1999; 99US-0150460.  
 PR 15-OCT-1999; 99US-0159849.  
 XX  
 XX (INCYTE) INCYTE GENOMICS INC.  
 PA Yue H, Tang YT, Bandman O, Hillman JT, Lai P, Au-Young J;  
 PI Reddy R, Yang J, Baughn MR, Lu DM, Azimzal Y, Patterson C;  
 DR WPI; 2001-091972/10.  
 XX P-PSDB; AAB68550.  
 PT New guanosine triphosphate-binding associated proteins (GBAP) and their  
 PT encoding nucleic acids, useful for treating and/or diagnosing diseases

PT associated with GBAP expression, such as cancer, diabetes and asthma -  
 XX  
 PS Claim 5; Pages 220-221; 233pp; English.

CC The present invention relates to novel human guanosine triphosphate  
 CC (GTP)-binding associated proteins (GBAPs; AAB6501-AAB6566) and their  
 CC coding sequences (AAF58301-AAF58366). The proteins and coding sequences  
 CC of the present invention are useful for treating a variety of disorders  
 CC including inflammation, AIDS, Addison's disease, anaemia,  
 CC arteriosclerosis, asthma, autoimmune disorders, grave's disease,  
 CC hepatitis, multiple sclerosis, cancer, diabetes, osteoporosis and  
 CC psoriasis.

CC  
 XX Sequence 1286 BP; 296 A; 355 C; 337 G; 298 T; 0 other;

Query Match 37.2%; Score 222; DB 22; Length 1286;  
 Best Local Similarity 62.2%; Pred. No. 5.6e-34;  
 Matches 367; Conservative 0; Mismatches 220; Indels 3; Gaps 1;

QY 11 AGACTAACGATTACCGCTGGTGGTGGGCGGCGGCGGCAAGAGCTGCTGG 70  
 D 456 AGATCAAGAGATTACCGCTGGTGGTGGGCGGCGGCGGCAAGAGCTGCTGG 515  
 QY 71 TGTGCGCTTGTGAAGGCGACGTTCCGACACCTACATCCCAACATCGACACT 130  
 D 516 TGCACAAAGTGGCGAGCGGCAACTTCCGTACATGAGTACCTCCGACCATGAAATACCT 575  
 QY 131 ACCGCGAGCTGATCAGCTGGCGCAACAAGAGCTGACCGCTCAGTACACACACCCAG 190  
 D 576 ACTCCAGATTCTGGGCTGAGCGACCGAGTGGCTTTCCCTCAGCTACCCAGACAGCAAGA 635  
 QY 191 GCAGCGACCAAGTCCCGGCGCATGACAGCGCTGTCATCCCAAGGCGCACCTTCATCC 250  
 D 636 GTGGCGAAGGCAACCGGCTCTGCAGCGCCACGTTATATAGCCCGGCGCACCGCTTCGCC 695  
 QY 251 TGTGTTCTCCGTCACGACGACAGCATGCTGGAGAGCTGGGCGCATCTACACAGCTCA 310  
 D 696 TGGTCTACTCAGTCAACCAAGAGAAACCTGGAAGAGCTGAAGGCTTCTATGAGCTGA 755  
 QY 311 TCGTGCGATCAAGGG---CAGCGTGGAGAGCATCCCGGTATGCTGGGCAACAGT 367  
 D 756 TCTCGAAGATTAAGATTAACCTGCTAATAGTTCCATGCTGTGGGCAATTAATA 815  
 QY 368 GCGATGAGACGACGAGGAGGTGACACGCGCGAGCGCGAGCGGTCGCCAAGAGTGA 427  
 D 816 GTGATGACACCCACCGGAGGTGGCGCTGATGATGATGATGATGATGATGATGATGATG 875  
 QY 428 AGTGGCTTTATGAGACCTCGGCGCAAGATGAACTACACGTCAGGAGCTTCTTCAGG 487  
 D 876 ATTGCGCTTCATGAGATTTTCAAGCAAGACCGATGTAATGTGACAGAGCTGTTCACA 935  
 QY 488 AGTGTGAGCGCTGGAGCGGCGGGAACATGAGCGCTCAACATGACGAGGCAAGCGCTCG 547  
 D 936 TGTGCTGTAATTACAAAGAAAAACCCACCGGCTTCAGAGAGCCGAGAAAGATCC 995  
 QY 548 GGAAGCAAGAGAGACAGACCGCTCAAGGCGCAATGCAACCTCATGTGA 597  
 D 996 AGATGCCCAACACCACTGAGAAAGCTGTTGACAAGTGCATATATATGGA 1045

RESULT 8  
 AAH15447  
 ID AAH15447 standard; cDNA; 1561 BP.

XX AAH15447;

XX 26-JUN-2001 (first entry)

DE Human cDNA sequence SEQ ID NO:13678.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

XX  
 PN EPI074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 XX 28-JUL-2000; 2000EP-0116126.  
 PF  
 XX 29-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX  
 DR WPI; 2001-318749/34.  
 XX  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 XX  
 PS Claim 8; SEQ ID 13678; 2537pp + CD ROM; English.  
 XX  
 CC The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to  
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 XX  
 XX Sequence 1561 BP; 422 A; 360 C; 373 G; 406 T; 0 other;

Query Match 37.2%; Score 222; DB 22; Length 1561;  
 Best Local Similarity 62.2%; Pred. No. 5.7e-34;  
 Matches 367; Conservative 0; Mismatches 220; Indels 3; Gaps 1;

QY 11 AGACTAACGATTACCGCTGGTGGTGGGCGGCGGCGGCAAGAGCTGCTGG 70  
 D 456 AGATCAAGAGATTACCGCTGGTGGTGGGCGGCGGCGGCAAGAGCTGCTGG 515  
 QY 71 TGTGCGCTTGTGAAGGCGACGTTCCGCGACCTACATCCCAACATCGAGACACT 130  
 D 516 TGCACAAAGTGGCGAGCGGCAACTTCCGTACATGATTAAGTTCGCGCAATGAAATACCT 624  
 QY 131 ACCGCGAGTGTATGCTGCGCAAGAGCGTGTGACACCTGACATGACAGACACACCG 190  
 D 625 ACTGCGAGTTGCTGGGCTGCGAGCCAGCGTGTCTTCCCTGACATCAGCAGACAGAGA 684  
 QY 191 GCAGCGACCAAGTCCCGGCGCATGACAGCGCTGTCCATCTCCCAAGGCGCAGCGCTTCATCC 250  
 D 685 GTGGGAGCGCAACCGCGCTCTGCGAGCGCCACGTTATATAGCGGCGGCGCGCTTCGTC 744  
 QY 251 TGTGTTCTCCGTCACGACGACGAGCATGCTGGAGAGCTGGGCGCATCTACACAGCTCA 310

```
Db 745 TGGTCTACTAGTCACCAAGAAACCTTGAAGAGCTGAAGGCTTATGAGCTGA 804
Oy 311 TCGTGCAGCAAGG---CAGCGTGGAGACATCCCGTGATGCTCGGCAACAAGT 367
Db 805 TCTGCAAGATCAAGATCAACACCTGCATTAAGTCCCATCGTGTGGGCAATTA 864
Oy 368 GCGATGAGACGACGCGGAGGTGACACCGCGAGCGGCGGTGGCCCAAGAGTGA 427
Db 865 GTATATACACCCACCGGAGGTGGCCCTGAATGATGTGGCCACCTGCGATGAGTGA 924
Oy 428 AGTGGCTTTCATGAGACCTGGCCCAAGTGAACATCAACAGCTGACCTCTCCAG 487
Db 925 ATTGGCGCTTTCATGAGATTTTCAGCCAGACGATGTAATGTGACGACCTGTCC 984
Oy 488 AGCTGCTGACGTGAGACGCGCGGACATGACCTCAACATGACGCGCAACGCTCG 547
Db 985 TCGTGTGAATTAAGAAAGAAACCCACCGCGCTCCGAGAGCCGCAAGAAATCC 1044
Oy 548 GGAAGCAGAAAGAGACAGACCGCGCTCAAGGCGCAATGCAACCTTCATGTGA 597
Db 1045 AGATGCCCAACACCACTGAGAGCTGCTTGAACAAGTGCAATATCATGTGA 1094
```

## RESULT 9

AAS87856 standard; cDNA: 3403 BP.

AAS87856;

13-FEB-2002 (first entry)

DNA encoding novel human diagnostic protein #23660.

Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder; ss.

Homo sapiens.

WO200175067-A2.

11-OCT-2001.

30-MAR-2001; 2001MO-US08631.

31-MAR-2000; 2000US-0540217.

23-AUG-2000; 2000US-0649167.

(HYSE-) HYSEQ INC.

Drmanac RT, Liu C, Tang YT;

WPI: 2001-639362/73.

P-PSDB: ABG23669.

New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.

Claim 1: SEQ ID NO 23660; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating

disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human CC diagnostic coding sequences of the invention. CC Note: The sequence data for this patent did not appear in the printed CC specification, but was obtained in electronic format directly from WIPO at ftp://wipo.int/pub/published\_pct\_sequences.

SQ Sequence 3403 BP; 909 A; 850 C; 887 G; 757 T; 0 other;

Query Match 37.2%; Score 222; DB 23; Length 3403;

Best Local Similarity 62.2%; Pred. No. 5,9e-34;

Matches 367; Conservative 0; Mismatches 220; Indels 3; Gaps 1;

```
Oy 11 AGATCAACGTTACCGCGTGTGTGTTGCGGGCGGCGGCTGGGCAAGACTCGCTGG 70
Db 2814 AGATCAAGATTAACCGCGTGTGTGTTGCGCACCGCTGTGTGGGAAAGTACCTGC 2873
Oy 71 TCGTGGCTTCGTGAAGGCGACGTTCCGAGACACTATACATCCACATGAGAGACCT 130
Db 2874 TGCACAAGTGGGCGAGCGCAACTTCCGTATAGTACCTGCGACATTGAAATACCT 2933
Oy 131 ACCGCGAGGTGATCAGTCGCGACAGACGCTGTGACGCTGCAATCAGACACCAACCG 190
Db 2934 ACTGCCAGTTGTGGGCTGCGACGAGTGCTCTTCCGTGCATACCGACGACGACAGA 2993
Oy 191 GCAGCCACGAGTTCGGGCGCATGCGGCGCTGTCATCTCCAGAGGCGCGGCTTATCC 250
Db 2994 GTGGCGAGCGCAACCGGCTGTGACGCGCATTTATAGCCGCGGCGCACGCTTGTGTC 3053
Oy 251 TGGTGTTCCTCGTACACGACAGTCTGAGAGAGTGGGCGCCATCTACACCTCA 310
Db 3054 TGGTCTACTAGTCACCAAGAAAGAAACCTGGAAGAGCGCTTATGAGACTGA 3113
Oy 311 TCGTGCAGATCAAGG---CAGCGTGAAGGACATCCCGTGATGCTGTGGCAACAAGT 367
Db 3114 TCTGCAAGATCAAGGTAACACCTCATTAAGTCCCATCGCTGTGGGCAATTA 3173
Oy 368 GCGATGAGACGCGGAGGAGTGACACGCGGCGGAGGCGGAGGCGCCACAGAGTGA 427
Db 3174 GTATGACACCCACCGGAGAGTGGCCCTGATGATGTGTCACCTGTGAGTGAAGTGA 3233
Oy 428 AGTGGCTTTCATGAGACCTCGGCCAAGATGAACAAACGTCAAGAGAGCTTTCAG 487
Db 3234 ATTGGCGCTTCATGAGATTTTCAGCCAGACGATGTAATGTGAGAGCTGTCCACA 3293
Oy 488 AGCTGCTGACGCTGAGAGCGCGCGGACATGAGCTCAACATGACGCGCAACGCTCG 547
Db 3294 TCGTGTGAATTAAGAAAGAAACCCACCAACCGCTCCAGAGGCGGAAAGAAATCC 3353
Oy 548 GGAAGCAGAAAGAGACAGACCGCGTCAAGGCGCAATGCAACCTTCATGTGA 597
Db 3354 AGATGCCCAACACCACTGAGAGAGCTGCTTGAACAAGTGCAATATCATGTGA 3403
```

## RESULT 10

AAV60578 standard; cDNA: 7397 BP.

AAV60578;

18-JAN-1999 (first entry)

Human tumour suppressor gene NOEY2.

NOEY2; tumour suppressor; human; breast cancer; ovary cancer; diagnosis; gene therapy; transgenic animal; ss.

Homo sapiens.

FH Key Location/Qualifiers  
 FT CDS 5468..6158  
 XX /\*tag= a  
 PN MO9842830-A2.  
 PD 01-OCT-1998.  
 XX 20-MAR-1998; 98WO-US05723.  
 PF 13-JAN-1998; 98US-0071263.  
 PR 21-MAR-1997; 97US-0041580.  
 XX (TEXA ) UNIV TEXAS SYSTEM.  
 PA Bast RC, Xu F, Yu Y;  
 PI WPI; 1998-532005/45.  
 DR P-PSDB; AAW/9898.  
 XX  
 XX New nucleic acid encoding NOEY2 tumour suppressor from ovarian  
 PT epithelium - useful for, e.g. treatment, diagnosis and prognosis of  
 cancer, particularly cancer of ovary and breast  
 PS Claim 4; Page 152-158; 182pp; English.  
 XX  
 XX This is the NOEY2 gene nucleotide sequence. NOEY2 is a novel human  
 CC tumour suppressor gene that is expressed in normal ovarian and  
 CC breast surface epithelial cells but is consistently absent or  
 CC down-regulated in ovarian and breast cancer cells. The open  
 CC reading frame encodes a 26 kDa polypeptide (see AAW/9898) that shows  
 CC homology to members of the Ras and Rap family. The gene is  
 CC isolated from human chromosome 1p31. A full-length cDNA clone  
 CC (see AAV60577) is also claimed. NOEY2 nucleic acids can be used in  
 CC the recombinant production of NOEY2 polypeptides, in the gene  
 CC therapy of cancer (especially ovarian and breast-related), and to  
 CC prepare transgenic animals useful as models of cancer. Fragments  
 CC of NOEY2 nucleic acids are also used as probes and primers for  
 CC diagnosis, including detection of mutations, or as antisense  
 CC molecules or ribozymes for reducing/eliminating NOEY2 activity. The  
 CC nucleic acids can also be used to screen for antitumour agents that  
 CC stimulate NOEY2, overcome lack of this protein or block expression  
 CC of mutant NOEY2.  
 XX  
 SQ Sequence 7397 BP; 2043 A; 1542 C; 1651 G; 2063 T; 98 other;  
 Query Match 36.9%; Score 220.4; DB 19; Length 7397;  
 Best Local Similarity 62.0%; Pred. No. 1.3e-33;  
 Matches 366; Conservative 0; Mismatches 221; Indels 3; Gaps 1;  
 11 AGAGTAACGATTAACCGGTGGTGGTGGGGGGGGGGGGGCAAGAGCTCGCTGG 70  
 5568 AGATCAGAGATTACCGGCTGCTAGTCCGACCCGCTGGTGGGAAAGTACGGCTGC 5627  
 71 TGTGCGCTTCGTGAGGGGCACTTCCGCGACACTTACATCCCAACATCAGAGACACT 130  
 5628 TGCACAGATGGGGGCAACTCCGTATAGTAAGTACCTCCACCACTTAATAATACCT 5687  
 131 ACCGGCAGGTGATCAGCTCGCAGAGAGCGTGTGCACGCTGCAGATCAGACACCACCG 190  
 5688 ACTGCCAGTTGCTGGCTGAGCCAGCGTGTGCTTCCCTGCATCACCAGACGACGA 5747  
 191 GCAAGCCACGATTCGCCGCGCATGACAGCGCTGTCCATCTCCAAAGGCGCAGCCCTTCATCC 250  
 5748 GTGGCGAGCGCAACCGGCTCTGCGAGCGCCAGTTATAGCCCGGGGCGACGCTTCGCGCC 5807  
 251 TGGTGTCTCCGTCACAGCAGATCGCTGAGAGAGCTGGGCGCCATTCAGACATCA 310  
 5808 TGGTCTACTAGTACCAAGAGAAACCCGTGGAAGACCTTAAGCCCTTCTATAGCTGA 5867  
 311 TCGTGCAGATCAAGGG---CAGCGTGAGAGACATCCCGGTGATGCTCGTGGGCAACAGT 367  
 5868 TCTGCAGATCAAGAGTAACAACTGCATAGTTCCCATGCTGCTGGTGGCAATAA 5927

368 GCGATGAGACCGAGCGGAGGTGACACAGCCGCGAGGCGCGGATGCGCCCAAGAGTGA 427  
 5928 GTGATGACACCCACCGGAGGTGCGCCCTGATGATGTCGCCACCTGTCGATGGAGTGA 5987  
 428 AGTGGCTTCATGAGAGACTCGGCGCAAGATGAACACTACAGTCAGAGAGCTTTCAGG 487  
 5988 ATTGGGCTTCATGAGAGATTTACGCTAAGACCGATGATGATGTGAGAGAGTGTTCACA 6047  
 488 AGTGTCTACGCTGGAGAGCGCGGAGAACATGAGCTTCAACATGAGCGGCAAGCGCTCG 547  
 6048 TGTGCTGAATTACAAAGAAAAAGCCACCGGCTTCAGAGAGCCCAAGAAATCC 6107  
 548 GGAAGCAGAGAGAGACAGCCGCTCAAGGCGCAATGCAACCTCATGTGA 597  
 6108 AGATGCCACACCACTGAGAAAGCTTGGACAAAGTGCATATCATGTGA 6157  
 RESULT 11  
 AAV60577  
 ID AAV60577 standard; cDNA; 1481 BP.  
 AC AAV60577;  
 DT 18-JAN-1999 (first entry)  
 XX  
 DE Human tumour suppressor gene NOEY2 cDNA.  
 XX  
 XX NOEY2; tumour suppressor; human; breast cancer; ovary cancer;  
 KW diagnosis; gene therapy; transgenic animal; ss.  
 XX Homo sapiens.  
 FH Key Location/Qualifiers  
 FT CDS 150..836  
 XX /\*tag= a  
 PN MO9842830-A2.  
 PD 01-OCT-1998.  
 XX 20-MAR-1998; 98WO-US05723.  
 PF 13-JAN-1998; 98US-0071263.  
 PR 21-MAR-1997; 97US-0041580.  
 XX (TEXA ) UNIV TEXAS SYSTEM.  
 PA Bast RC, Xu F, Yu Y;  
 PI WPI; 1998-532005/45.  
 DR P-PSDB; AAW/9898.  
 XX  
 XX New nucleic acid encoding NOEY2 tumour suppressor from ovarian  
 PT epithelium - useful for, e.g. treatment, diagnosis and prognosis of  
 cancer, particularly cancer of ovary and breast  
 PS Claim 4; Page 148-150; 182pp; English.  
 XX  
 XX This is the complete NOEY2 cDNA nucleotide sequence. NOEY2 is a  
 CC novel human tumour suppressor gene that is expressed in normal  
 CC ovarian and breast surface epithelial cells but is consistently  
 CC absent or down-regulated in ovarian and breast cancer cells.  
 CC The open reading frame encodes a 26 kDa polypeptide (see AAW/9898)  
 CC that shows homology to members of the Ras and Rap family.  
 CC Differential display PCR was used to isolate and clone the novel  
 CC gene, and RACE was used to amplify the full-length NOEY2 cDNA from  
 CC human ovarian surface epithelial cells. A NOEY2 genomic DNA  
 CC sequence (see AAV60578), isolated from human chromosome 1p31, is also  
 CC claimed. NOEY2 nucleic acids can be used in the recombinant  
 CC production of NOEY2 polypeptides, in the gene therapy of cancer  
 CC (especially ovarian and breast-related cancers), and to prepare  
 CC transgenic animals useful as models of cancer. Fragments of NOEY2









PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249219.  
PR 17-NOV-2000; 2000US-0249220.  
PR 17-NOV-2000; 2000US-0249221.  
PR 17-NOV-2000; 2000US-0249222.  
PR 17-NOV-2000; 2000US-0249223.  
PR 17-NOV-2000; 2000US-0249224.  
PR 17-NOV-2000; 2000US-0249225.  
PR 17-NOV-2000; 2000US-0249226.  
PR 17-NOV-2000; 2000US-0249227.  
PR 17-NOV-2000; 2000US-0249228.  
PR 17-NOV-2000; 2000US-0249229.  
PR 17-NOV-2000; 2000US-0249230.  
PR 17-NOV-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251858.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251889.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Barash SC, Ruben SM;

DR WPI; 2001-483426/52.

PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating cancers and  
PT metastasis -

PS Disclosure; SEQ ID NO 36047; 3071bp + Sequence Listing: English.

XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
XX amino acid sequences given in AAK62170 to AAK61921. (I) have cytosolic  
XX activity, and can be used in gene therapy and vaccine production. (I)  
XX proteins and polynucleotides may be used in the prevention, diagnosis and  
XX treatment of diseases associated with inappropriate (I) expression. For  
XX example, they may be used to treat disorders associated with decreased  
XX expression by rectifying mutations or deletions in a patient's genome  
XX that affect the activity of (I) by expressing inactive proteins or to  
XX supplement the patient's own production of (I). Additionally, (I)  
XX polynucleotides may be used to produce the secreted (I), by inserting  
XX the nucleic acids into a host cell and culturing the cell to express the  
XX protein. (I) proteins and polynucleotides may be used to prevent,  
XX diagnose and treat immune/hematopoietic-related diseases, especially  
XX cancers and cancer metastases of hematopoietic-derived cells. AAK64703  
XX to AAK67694 represent human immune/hematopoietic antigen genomic  
XX sequences from the present invention. AAK54942 to AAK54950 and AAK62169  
XX represent sequences used in the exemplification of the present invention.

SO Sequence 4933 BP; 1544 A; 1175 C; 985 G; 1229 T; 0 other;

Query Match 29.4%; Score 175.4; DB 22; Length 4933;  
Best Local Similarity 58.2%; Pred. No. 5.5e-25;

Matches 308; Conservative 0; Mismatches 221; Indels 0; Gaps 0;

QY 19 GATTACCGGCTGTGTTTCGGGGGCGGCGGCGGCAAGACCTGCTGCTGCC 78  
DB 4469 GAGTACAAAGTGTGTGCTGCTGGGCGGCGGCGGCAAGCTCCGCTCACCGTGCAG 4410  
QY 79 TTTCGTAAGGAGCGAGCTTCCGCACTCATCCACCATGAGAGACATCAACCGGAG 138  
DB 4409 TTTCGTAAGGAGCGAGCTTCCGCACTCATCCACCATGAGAGACATCAACCGGAG 4350  
QY 139 GTGATCAGCTGCGACAAAGAGCTGTGCAAGCTGACATGACAGACACCGGACCGAC 198  
DB 4349 GAGATTGAGAGTGTGAGCTGTGCGCGCTGTGAGAGATCTGTGATGCGGCGGACCGAG 4290

QY 199 CAGTCCCGGCGCATGACGCGCTGTCTCATCTCCAGGCGCCAGCCCTTCACTGTGTTTC 258  
DB 4289 CAGTCCCGGCGCATGACGCGCGCTGTCTCATCTCCAGGCGCCAGCCCTTCACTGTGTTTC 4230  
QY 259 TCCGTACACAGCAAGAGCTGTGAGAGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAG 318  
DB 4229 AGCTCTGTCAACGAG 4170  
QY 319 ATCAAGGAG 378  
DB 4169 GTGACAGGAG 4110  
QY 379 CAGCGGAG 438  
DB 4109 GAGCGGAG 4050  
QY 439 ATGAG 498  
DB 4049 ATGAG 3990  
QY 499 CTGAG 547  
DB 3989 CAGATGAG 3941

RESULT 14

AAK81234/c  
ID AAK81234 standard; DNA; 4934 BP.

XX AAK81234;

XX 07-NOV-2001 (first entry)

DE Human immune/hematopoietic antigen genomic sequence SEQ ID NO:36046.

KW Human; immune; hematopoietic; immune/hematopoietic antigen; cancer;

KW cytostatic; gene therapy; vaccine; metastasis; ds.

XX Homo sapiens.

XX WO200157182-A2.

XX 09-AUG-2001.

XX 17-JAN-2001; 2001WO-US01354.

XX 31-JAN-2000; 2000US-0179065.

XX 04-FEB-2000; 2000US-0180628.

XX 24-FEB-2000; 2000US-0184664.

XX 02-MAR-2000; 2000US-0186350.

XX 16-MAR-2000; 2000US-0188874.

XX 17-MAR-2000; 2000US-0190076.

XX 18-APR-2000; 2000US-0198123.

XX 19-MAY-2000; 2000US-0205515.

XX 07-JUN-2000; 2000US-0209467.

XX 28-JUN-2000; 2000US-0214886.

XX 30-JUN-2000; 2000US-0215135.

XX 07-JUL-2000; 2000US-0216447.

XX 07-JUL-2000; 2000US-0216880.

XX 11-JUL-2000; 2000US-0217487.

XX 11-JUL-2000; 2000US-0217496.

XX 14-JUL-2000; 2000US-0218290.

XX 26-JUL-2000; 2000US-0220963.

XX 26-JUL-2000; 2000US-0220964.

XX 14-AUG-2000; 2000US-0224518.

XX 14-AUG-2000; 2000US-0224519.

XX 14-AUG-2000; 2000US-0225213.

XX 14-AUG-2000; 2000US-0225214.

XX 14-AUG-2000; 2000US-0225266.

XX 14-AUG-2000; 2000US-0225267.

XX 14-AUG-2000; 2000US-0225268.

XX 14-AUG-2000; 2000US-0225270.

XX 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 23-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.

PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249219.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 06-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.

## (HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-483426/52.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
useful for preventing, diagnosing and/or treating cancers and  
metastasis -

Disclosure; SEQ ID NO 36046; 3071pp + Sequence Listing; English.

AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
amino acid sequences given in AAM82170 to AAM91921. (I) have cytosolic  
activity, and can be used in gene therapy and vaccine production. (I)  
proteins and polynucleotides may be used in the prevention, diagnosis and  
treatment of diseases associated with inappropriate (I) expression. For  
example, they may be used to treat disorders associated with decreased  
expression by rectifying mutations or deletions in a patient's genome  
that affect the activity of (I) by expressing inactive proteins or to  
supplement the patients own production of (I). Additionally, (I)  
polynucleotides may be used to produce the secreted (I), by inserting  
the nucleic acids into a host cell and culturing the cell to express the  
protein. (I) proteins and polynucleotides may be used to prevent,  
diagnose and treat immune/hematopoietic-related diseases, especially  
cancers and cancer metastases of hematopoietic-derived cells. AAK64703  
to AAK87694 represent human immune/hematopoietic antigen genomic  
sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
represent sequences used in the exemplification of the present invention.

Sequence 4934 BP; 1542 A; 1175 C; 985 G; 1232 T; 0 other;

Query Match 29.4%; Score 175.4; DB 22; Length 4934;  
Best Local Similarity 58.2%; Pred. No. 5; Se-25;  
Matches 308; Conservative 0; Mismatches 221; Indels 0; Gaps 0;

```

QY 19 GATTACCGCGTGGTGTGTTGCGGGCGCGGCGGCAAGAGCTGCTGCTGCGC 78
DB 4470 GAGTACAAAGTGGTGGTGGCTGGGGCGGCGGCAAGTCCGGCTCACCCTGAC 4411
QY 79 TTGCTGAAGGCGACGTTCCCGACACCTACATCCCACTGAGAGCACTACCGGAC 138
DB 4410 TTGCTGAAGGCGCTCTTCATCAGAGTACGACCCGACCATTCGAACTTTACCGGAC 4351
QY 139 GTGATCAGCTGCGCAAGAGCGTGTGACGCTGACATCCATCCACACACCAGGAC 198
DB 4350 GAGATTGAGGTGACTGCTGCGCGCTGCTGCTGAGATCCTGATACGGCGGACCGAC 4291
QY 199 CAGTTCCTCCGCGCATGACGCGCTGTCCATCTCCAAAGGCGCCCTTCATCTGATGTT 258
DB 4290 CAGTTCGCGCTCATGCGGACCTGTACATCAAGACGGCGGCTTCATCTGCTGTTAC 4231
QY 259 TCCGTACCAAGCAAGAGTGTGAGAGAGCTGGGCGCCATCTACAACTCATCTGTCAG 318
DB 4230 AGCTCGTCAACCAAGAGAGCTTCAGAGATCAAGCCCATGCGGACCATCATCTCGC 4171
QY 319 ATCAAGGCGACGCTGAGAGACATCCCGTGTGATGCTGTGGCAACAAGTCCGATGAGAC 378
DB 4170 GTCAACCGGTACGACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4111
QY 379 CAGCGGAGGTGACACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 438
DB 4110 GAGCGGAGGTGCTCTGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4051
QY 439 ATGAGAGCGTGGCGCAAGTGTACAACTACAACTGCAAGAGCTCTTCCAGAGAGCTGAC 498
DB 4050 ATGAGAGCGTGGCGCAAGTGTACAACTGCAAGAGCTCTTCCAGAGAGCTGAC 3991
QY 499 CTGAGAGCGCGCGGCAAGTGTACAACTGCAAGAGCTCTTCCAGAGAGCTGAC 547
DB 3990 CAGATGACTACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3942

RESULT 15
AA003212
ID AA003212 standard; DNA; 558 BP.
AC AA003212;
DT 10-JUL-1990 (first entry)
DE RAP2 Gene encoding at least one peptide associated with ras oncogene.
KW ras oncogene; cancer; GTPase; ds.
XX Synthetic.
FH Key Location/Qualifiers
FT CDS 4..552
FT /tag= a
XX WO9000179-A.
XX PN 11-JAN-1990.
XX PD 11-JAN-1990.
XX PE 04-JUL-1989; 89WO-FR00348.
XX PR 04-JUL-1988; 88FR-0009031.
XX PA (INRM ) INSEHM INST NAT SANTE.
XX PI Tavitlan A, Pizon V, Chardin P;
XX DR WPI; 1990-037122/05.
XX DR P-PSDB; AAR05076.
XX PT DNA encoding sequences associated with human oncogenes -
XX and derived antibodies, useful for in vitro diagnosis of cancer
XX and for therapy.

```

```

XX PS Disclosure; Fig 3a-b; 92pp; French.
XX CC Derived peptides can fix GTP and GDP and have GTPase activity. Abs raised
XX CC to these peptides can diagnose associated diseases - absence or excess of
XX CC the product indicates loss of balance between anti- and oncogenic factors
XX CC in the cells. The product may be used to treat conditions associated with
XX CC ras or rap genes.
SQ Sequence 558 BP; 142 A; 143 C; 165 G; 108 T; 0 other;
Query Match 27.9%; Score 166.4; DB 11; Length 558;
Best Local Similarity 58.8%; Pred. No. 2.7e-23;
Matches 287; Conservative 0; Mismatches 201; Indels 0; Gaps 0;
QY 18 CGATTACCGCGTGGTGTGTTGCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 77
DB 9 CGAGTACAAAGTGGTGGTGGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 68
QY 78 CTTCGTGAAGGCGACGTTCCGCGACACTTACATCCCACTGAGAGCACTACCGGCA 137
DB 69 GTTCGTGACCGGCACTTCATCGAAGAAATACGACCCCATCGAGACTTCTTACCGCAA 128
QY 138 GTGATCAGCTGCGACAGAGCGTGTGACGCTGACAGATCAGACACCAACCGGCAAGCA 197
DB 129 GGAATCAGAGTGGATTCGTCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 188
QY 198 CAGATTCCCGGCGCATGACGCGCTGTTCATCTTCAAGGCGCACGCTTCATCTGCTGTT 257
DB 189 GCAGTTGCGCTCATGCGGACCTGTATCAATCAAGAACGCGGCGGCTTCATCTGCTGCT 248
QY 258 CTTCGTGACAGCAAGCAGAGCGTGTGAGAGAGCGTGGGCGGCGGCGGCGGCGGCGG 317
DB 249 CAGCTCTGCAACCAAGAGCTTCCAGAGACATCAAGCCCATGCGGAGCGAGATCATCCG 308
QY 318 GATCAAGGCGCGGTGAGAGACATCCCGGATGCTGTGTGGCAACAAGTCCGATGAGAC 377
DB 309 CGTGAAGCGGTATGAGAAAGTGCACATCTTGTGTGGAAACAAGTGCACCTGGAAGA 368
QY 378 GCAGCGGAGGTGACACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 437
DB 369 TGAGAGAGAGATGCTGCGACGCAAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 428
QY 438 CATGAGACCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 497
DB 429 TATGAACTTCGCTTAAGGTAAGAAATGTTGAGCACTTCTTGGCAAAATGTTGAG 488
QY 498 GCTGAGAG 505
DB 489 GCAGATGA 496

```

Search completed: March 23, 2003, 15:20:19  
Job time : 709.666 secs

GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 23, 2003, 14:20:51 ; Search time 3881.38 seconds  
(without alignments)  
2491.051 Million cell updates/sec

Title: US-09-873-546-4

Sequence: 1 atcgccgacacagatgtaacga.....gcaaatgcacctctatgta 597

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

otal number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Listing first 45 summaries

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estlin:\*

4: em\_estinu:\*

5: em\_estrov:\*

6: em\_estrpl:\*

7: em\_estro:\*

8: em\_hc:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_hc:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estrom:\*

17: gb\_gss:\*

18: em\_gss\_hum:\*

19: em\_gss\_inv:\*

20: em\_gss\_pin:\*

21: em\_gss\_vrt:\*

22: em\_gss\_fun:\*

23: em\_gss\_mam:\*

24: em\_gss\_mus:\*

25: em\_gss\_other:\*

26: em\_gss\_pro:\*

27: em\_gss\_rtd:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	525.4	88.0	664	12	BG699685 602681538
2	470	78.7	662	10	BB648303 BB648303
3	454.6	76.1	838	14	BO887987 BO887987
4	416.4	69.7	470	9	AI497811 AI497811
5	402.4	67.4	454	9	AA723009 AA723009
6	386.6	64.8	458	10	BE095820 BE095820

Result No.	Score	Query Match	Length	DB ID	Description
7	381.4	63.9	456	10	BE095936 BE095936
8	353.8	59.3	422	9	AI764459 AI764459
9	349.2	58.5	419	10	AM533412 AM533412
10	339.4	56.9	801	12	BG702913 BG702913
11	336.6	56.4	645	13	BI547693 BI547693
12	328.4	55.0	702	10	AM050720 AM050720
13	322.2	54.0	786	13	BI552244 BI552244
14	312.2	52.3	628	13	BM172422 BM172422
15	311.2	52.1	709	12	BG699614 BG699614
16	302.2	50.6	632	13	BI475251 BI475251
17	302	50.5	626	10	AM134357 AM134357
18	301.6	50.5	618	9	AU171413 AU171413
19	294.2	49.3	893	14	BO922385 BO922385
20	281.8	47.2	466	10	BE019848 BE019848
21	265.6	44.5	562	12	BF193212 BF193212
22	261.8	43.9	573	13	BM186599 BM186599
23	246.8	41.3	558	13	BM034832 BM034832
24	237.8	39.8	478	10	AM784881 AM784881
25	237.6	39.8	448	10	AM656952 AM656952
26	228.2	38.2	687	14	BM944010 BM944010
27	220.4	36.9	467	12	BG982366 BG982366
28	213.2	35.7	349	10	BB871148 BB871148
29	209.2	35.0	1141	13	BM544956 BM544956
30	204.2	34.2	756	12	BG707443 BG707443
31	203.8	34.1	847	9	AL668307 AL668307
32	203.4	34.1	1119	12	BE885242 BE885242
33	203.2	34.0	908	13	BI603847 BI603847
34	202.8	34.0	961	9	AL664187 AL664187
35	200.4	33.6	550	13	BM263033 BM263033
36	195.4	32.7	840	12	BF213048 BF213048
37	192.4	32.2	560	13	BI539761 BI539761
38	192.4	32.2	565	13	BM107153 BM107153
39	189.4	31.7	681	9	AL667679 AL667679
40	189.2	31.7	810	13	BI599064 BI599064
41	188.6	31.6	895	9	AL666801 AL666801
42	187	31.3	727	12	BG776910 BG776910
43	185.2	31.0	799	13	BI668300 BI668300
44	178.2	29.8	497	13	BI849738 BI849738
45	178.2	29.8	591	10	BE605570 BE605570

## ALIGNMENTS

RESULT 1

LOCUS BG699685 664 bp mRNA linear EST 07-MAY-2001

DEFINITION 602681538F1 NIH\_MGC\_95 Homo sapiens cDNA clone IMAGE:4814337 5', mRNA sequence.

ACCESSION BG699685

VERSION BG699685.1 GI:13968244

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 664)

AUTHORS NIH-MGC. <http://mgc.ncl.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabs-f@mail.nih.gov](mailto:cgabs-f@mail.nih.gov)  
Tissue procurement: Miklos Palokovits, M.D., Ph.D.  
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshituki and Piero Carninci (RIKEN)  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Plate: LLAM10709 row: k column: 10  
High quality sequence stop: 662.  
Location/Qualifiers

source  
1. .664  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image="4814337"  
/clone\_lib="NH\_MGC\_95"  
/tissue\_type="hippocampus"  
/lab\_host="DH10B"  
/note="Organ: brain; Vector: pBluescriptR (modified pBluescript KS+); Site\_1: BamHI; Site\_2: SalI-XhoI (gtcgag ) ; Oligo-dt primed using primer 5'-TTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.5 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the cap-trapper method (carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH-MGC Library."  
BASE COUNT 131 a 212 c 229 g 92 t  
ORIGIN

Query Match 88.0%; Score 525.4; DB 12; Length 664;  
Best Local Similarity 99.8%; Pred. No. 7,6e-96;  
Matches 526; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 ATCCGGAGACAGATTAACGATTAACCGCGTGTGTGTGGGGGGGGCGGCGGAG 60  
138 ATCCGGAGACAGATTAACGATTAACCGCGTGTGTGTGGGGGGGGCGGCGGAG 197  
61 ACCTGCTGT 120  
198 ACCTGCTGT 257  
121 GAGGACACCTACCGGAGGTGATGAGTGTGTGTGTGTGTGTGTGTGTGTGTGT 180  
256 GAGGACACCTACCGGAGGTGATGAGTGTGTGTGTGTGTGTGTGTGTGTGTGT 317  
181 GACACACCTACCGGAGGTGATGAGTGTGTGTGTGTGTGTGTGTGTGTGTGT 240  
318 GACACACCTACCGGAGGTGATGAGTGTGTGTGTGTGTGTGTGTGTGTGTGT 377  
241 GCCTTCATCCTGT 300  
378 GCCTTCATCCTGT 437  
301 TACAAGCTCATGCTGATCAATCAAGGAGCGGTGAGGACATCCCGTGTGTGTGT 360  
438 TACAAGCTCATGCTGATCAATCAAGGAGCGGTGAGGACATCCCGTGTGTGTGT 497  
361 AACAAAGTCGATGAGAGCGGAGGAGGTGAGACGCGGAGCGGAGCGGAGCGCA 420  
498 AACAAAGTCGATGAGAGCGGAGGAGGTGAGACGCGGAGCGGAGCGGAGCGCA 557  
421 GAGTGAAGTGGCTTTTCATGAGACCTTCGCGCAAGATGAACTACACGTCGAAG 480  
558 GAGTGAAGTGGCTTTTCATGAGACCTTCGCGCAAGATGAACTACACGTCGAAG 617  
481 TTCCGAGAGCTGCTGAGCGTGGAGCGCGCGGAACATGAGCGCTCAA 527  
618 TTCCGAGAGCTGCTGAGCGTGGAGCGCGCGGAACATGAGCGCTCAA 664

RESULT 2  
BB648303  
LOCUS BB648303 RIKEN full-length enriched, 16 days embryo head Mus 662 bp mRNA linear EST 26-OCT-2001  
DEFINITION BB648303 RIKEN full-length enriched, 16 days embryo head Mus  
ACCESSION BB648303  
VERSION BB648303.1 GI:16482559  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 662)

AUTHORS  
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shirai,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.  
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)  
Unpublished (2001)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gs.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/  
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
wagl,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watanabe,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamana,I., Alizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.  
Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)  
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.  
e mouse tissues.

FEATURES  
source  
1. .662  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone\_image="C130007J02"  
/clone\_lib="RIKEN full-length enriched, 16 days embryo head"  
/sex="mixed"  
/tissue\_type="head"  
/dev\_stage="16 days embryo"  
/lab\_host="DH10B"  
/note="Site\_1: SalI; Site\_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGATTCGAGTAAATTAATATCCGCCGCCGCC 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTCGAGTAAATTAATTAATCCGCCGCCGCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I"  
BASE COUNT 144 a 191 c 222 g 105 t  
ORIGIN

Tissue Procurement: Susan L. Sullivan, PhD.  
 cDNA Library Preparation: Regesen, Invitrogen Corp.  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL).  
 DNA Sequencing by: Agencourt Bioscience Corporation.  
 Clone distribution: MCC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLNL13878 row: g column: 12  
 High quality sequence stop: 399.

QY	1	ATGCGGAAACAGATAACGATTACCCGGTGGTGGTTCCTGGGGGCGGGCGGTGGGCAAG	50
Db	9	ATGCCAGAACAGACCAATGACTACCGGTGTGTCTGTTCCGGCGAGGGCGGTGGGCAAG	58
QY	61	AGCTCGGTGTGTCGTCCGCTTCGTGAAGGGCACGTTTCGCGACACCTACATCCCAACATC	120
Db	69	AGCTCGGTGTGTCCTCCGCTTGTGAAGGGGACGTTTGTGCACCTACATCCCAACATA	128
QY	121	GAGACACACTACCGGCGAGGTGATGAGCTGGCACAAGAGCGTGTGCACGCTGCAGATTCACA	180
Db	129	GAGAGACAGCTACCGGCGAGGTGATGAGTGTGACAAAGAGCGTGTGCACACATGCAATACG	188
QY	181	GACACACACCGGCACGCCACGATTTCCCGGCCATTCGACGCGCTGTCTATCTCCAAAGGCGAC	240
Db	189	GACACACACCGGCACGCCACGATTTCCCGGCCATTCGACGCGGTGTCTATCTCCAAAGGCGAC	248
QY	241	GCCCTTCATCTGTGTCTTCTCCGTCCAGCAGCAAGCAGTGCCTGTGAGAGAGCTGGGCCATC	300
Db	249	GCCCTTATCTGTGTCTTCTCCGTCCAGCAGCAAGCAGTGCCTGTGATGAGTGTGAGGCCATC	308
QY	301	TACCAAGCTCATCTGTGCAGATTAAGGGCAGCGTGGAGGACATCCCGTGATCTCTGTGGGC	360
Db	309	TACCAAGTTCATCTGTGCAGATTAAGGGGAGCGCTGGAGGACATCCCAATCATGTCTGTGGGG	368
QY	361	AACAAAGTCGATGTGAGACGACGCGGAGGTGACACGCGCGAGGGGCGAGCGGTGGGCCAA	420
Db	369	AACAAAGTGTGATGTGAGACGACGCGGAGGTGACACGCGCGAGGGGCGAGCGGTGGGCCAA	428
QY	421	GAGTGAAGTACGCGCTTTCATGTGAGAGCCTCGCGCCAAATGAACTTCAACGTCGAAGGAGCTC	480
Db	429	GAGTGAAGTACGCGCTTTCATGTGAGAAACCTCACGAAAGATGAACCTTACACGTGAAGGAGCTG	488
QY	481	TTCCAGAGAGCTGTGACGCTGTGAGACGCGCGCGGAACATGAGCCTTCACACATGAGCGGCAG	540
Db	489	TTCCAGAGAGCTGTGACGCTGTGAGACGCGCGCGGAACCTTCACACGCTGTGAGCGGCAGC	548
QY	541	CGCTCCGGGAGGCAAGGAGGACGACAGCGGTCTTCAAGGGCAAAATCCACCTCATGTGCA	597
Db	549	CGCTCCGGGAGGCAAGGAGGACGACAGCGGTCTTCAAGGGCAAAATCCACCTCATGTGCA	602

ACCESSION	AI497811	
VERSION	AI497811.1	GI:4389793
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi	
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	

Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,  
 Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
 Bonaldo, Ph.D.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
[www.bio.linnl.gov/bdrr/image/image.html](http://www.bio.linnl.gov/bdrr/image/image.html)  
 Insert length: 721 Std Error: 0.00  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 455.  
 Location/Qualifiers  
 1.470

Query Match	69.7%	Score 416.4;	DB 9;	Length 470;
Best Local Similarity	99.5%	Pred. No. 5.5e-74;		
Matches 417; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;

QY	1	ATGCGGAAACAGATTAACGATACCGCGTGGTGTTCGGGGCGGGCGGTGGGGCAAG	60
Db	51	ATGCGGAAACAGATTAACGATACCGCGTGGTGTTCGGGGCGGGCGGTGGGGCAAG	110
QY	61	AGCTCGGTGTGTCTGCGCTTCGTGAAGGGCAGCTTCGCGACACCTACATCCCCACATC	120
Db	111	AGCTCGGTGTGTCTGCGCTTCGTGAAGGGCAGCTTCGCGACACCTACATCCCCACATC	170
QY	121	GAGGACACCTACCGGCAAGTATCAGTCTGCCAACAAGACGTTGCAACGTGCAGATCACA	180
Db	171	GAGGACACCTACCGGCAAGTATCAGTCTGCCAACAAGACGTTGCAACGTGCAGATCACA	230
QY	181	GACACACACGGCAGGACACAGTCCCGGCCATATGCAGCGCTTGCATCTCCCAAGGGCCAC	240
Db	231	GACACACACGGCAGGACACAGTCCCGGCCATATGCAGCGCTTGCATCTCCCAAGGGCCAC	290
QY	241	GCTTCATCTGTGTGTTCCTCCGTCAACCAGACAGTCGCTGGAGAGCTGGGGCCATTC	300
Db	291	GCTTCATCTGTGTGTTCCTCCGTCAACCAGACAGTCGCTGGAGAGCTGGGGCCATTC	350
QY	301	TACAAAGCTATCTGTGCAGATCAAGGGCAGCGTTGAGAGCAATCCCGTGATGCTCGTGGGC	360
Db	351	TACAAAGCTATCTGTGCAGATCAAGGGCAGCGTTGAGAGCAATCCCGTGATGCTCGTGGGC	410
QY	361	AACAAAGTCGATGTGAGAGCGACGCGGGAAGGTGACACACGCCGAGCGCAGCGGTGGCCCA	419
Db	411	AACAAAGTCGATGTGAGAGCGACGCGGGAAGGTGACACACGCCGAGCGCAGCGGTGGCCCA	469

REFERENCE  
AUTHORS  
1 (bases 1 to 454)  
Hillier, L., Allen, M., Bowles, L., Dubnque, T., Getse, G., Jost, S., Kitzman, D., Kucuba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Marti, J., Moore, B., Schellenberg, K., Stepec, M., Tan, F., Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.  
TITLE  
JOURNAL  
WashU-NCI human EST project  
Unpublished (1997)  
COMMENT  
Contact: Wilson RK

Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: [est@atson.wustl.edu](mailto:est@atson.wustl.edu)  
This clone is available royalty-free through LENT; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Seq primer: -40mls fwd. Et from Amer sham  
High quality sequence stop: 455.

```

FEATURES
    SOURCE
        Location/Qualifiers
            1..454
                /organism="Homo sapiens"
                /db_xref="GDB:1305294"
                /db_xref="taxon:9606"
                /clone="IMAGE:413600"
                /clone_1fb="Soares.pineal.gland.N3Hpc"
                /lab_host="DH10B (ampicillin resistant)"
                /note="Organ: pineal gland; Vector: pT73D (Pharmacia)
                    with a modified polylinker. Site:1: Not I; Site:2: Eco RI;
                    1st strand cDNA was primed with a Not I - oligo(dT) primer
                    [5' TGTTACCATCTGTAAGTGGAGCGCCGCGCTTTTCTTTTCTTTTCTTTT
                    3']
                    , double-stranded cDNA was size selected, ligated to Eco
                    RI adapters (Pharmacia), digested with Not I and cloned
                    into the Not I and Eco RI sites of a modified pT73 vector
                    (Pharmacia). Library constructed by Bento Soares and
                    M.Palima Bonaldo."
            BASE COUNT
                92 a      148 c      148 g      66 t
                ORIGIN

```

Query Match	Best Local Similarity	67.48	Score 402.4	DB 9	Length 454
Matches 403	Conservative	99.88	Pred. No. 3.5e-71	Mismatches 1	Indels 0
				Gaps 0	
QY	1	ATGCGGAGAACAGATTAACGATTACCGCGTGGTGGTTCGGGGCGGGCGGTGGGCAAG	60		
Db	49	ATGCGGAGAACAGATTAACGATTACCGCGTGGTGGTTCGGGGCGGGCGGTGGGCAAG	108		
QY	61	AGCTGCGTGGTGTCTGCGCTTCGTGAAGGGCAAGTTCGCGCAACCTACATATCCCAACATC	120		
Db	109	AGCTGCGTGGTGTCTGCGCTTCGTGAAGGGCAAGTTCGCGCAACCTACATATCCCAACATC	168		
QY	121	GAGGACACACTTACCGGGCAGGTGATCAGCTGCGACAAAGACGTGTGCACGCTGCAGATCACA	180		
Db	169	GAGGACACACTTACCGGGCAGGTGATCAGCTGCGACAAAGACGTGTGCACGCTGCAGATCACA	228		
QY	181	GACACACACGGGCAAGCCACCAATTCGGGGCCATGCAAGCGCTGTCCATATCTCCAAGGGCCAC	240		
Db	229	GACACACACGGGCAAGCCACCAATTCGGGGCCATGCAAGCGCTGTCCATATCTCCAAGGGCCAC	288		
QY	241	GCCTCATCTCTGGGTCTTCGTCGTCACCAAGACAGTGCCTGGAGAGAGCTGGGGCCATC	300		
Db	289	GCCTCATCTCTGGGTCTTCGTCGTCACCAAGACAGTGCCTGGAGAGAGCTGGGGCCATC	348		



QY 301 TACAAGCTCATGTCGATCATCAAGGCGACGCTGGAGACATCCCGTATGCTGTGGC 360  
|||||  
Db 349 TACAAGCTCATGTCGATCATCAAGGCGACGCTGGAGACATCCCGTATGCTGTGGC 408  
QY 361 AACAGTCCGATGAGACGCGGAGGTGAGACACGCCGAGGC 404  
|||||  
Db 409 AACAGTCCGATGAGACGCGGAGGTGAGACACGCCGAGGC 452

RESULT 6  
BE095820/c 458 bp mRNA linear EST 12-JUN-2000  
LOCUS  
DEFINITION UI-R-BU0-a02-h-08-0-UI.s1 UI-R-BU0 Rattus norvegicus cDNA clone  
BE095820  
ACCESSION UI-R-BU0-a02-h-08-0-UI 3', mRNA sequence.  
VERSION  
KEYWORDS BE095820.1 GI:8486751  
EST.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 458)  
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
TITLE Normalization and subtraction: two approaches to facilitate gene  
discovery  
JOURNAL Genome Res. 6 (9), 791-806 (1996)  
MEDLINE 97044477  
COMMENT Contact: Soares, MB  
Program for Rat Gene Discovery and Mapping  
University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: msouares@blue.weeg.uiowa.edu  
Oligo-dT track not found, Not I site shown in beginning of sequence  
is likely internal to the message. cDNA library Preparation: M.B.  
Soares Lab Clone distribution: clones will be available through  
Research Genetics (www.resgen.com)  
Seq primer: M13 Forward  
POLYA-No.

FEATURES  
Source  
Location/Qualifiers  
1..458  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-BU0-a02-h-08-0-UI"  
/clone\_1lb="UI-R-BU0"  
/dev\_stage="adult"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pUT3D-Pac (Pharmacia) with a modified  
polylinker. Site\_1: Not I; Site\_2: Eco RI; The UI-R-BU0  
library is a subtracted library derived from a mixture of  
eye and ganglia tissues. For a detailed description of the  
library from which this clone was derived, please visit  
our web site at [ratest.eng.uiowa.edu](http://ratest.eng.uiowa.edu). The subtraction has  
been previously described in (Bonaldo, Lennon and Soares,  
Genome Research 6:791-806, 1996)  
TAG\_SEQ=None found"

BASE COUNT 67 a 142 c 140 g 109 t  
ORIGIN

Query Match 64.8%; Score 386.6; DB 10; Length 458;  
Best Local Similarity 90.4%; Pred. NO. 5.3e-68;  
Matches 413; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 107 ACATCCCAACCATGAGAGACCTTACCGGAGTATGATGCGGACAGAGCGTGTGCA 166  
|||||  
Db 458 ACATCCCAACCATGAGAGACCTTACCGGAGTATGATGCGGACAGAGCGTGTGCA 399  
|||||  
QY 167 CGGTGACATGACAGACACCAACCGGACGACCACTGCTCCGGCCATGAGCGCCGTGTCCA 226  
|||||  
Db 398 CACTGACATGACAGACACCAACCGGACGACCACTGCTCCGGCCATGAGCGCGGTGTCCA 339  
|||||

QY 227 TCTCCAAAGGGCAGCGCTTTCATCTCGTGTCTCTCCGTCACGACGAGTGTGTGAGG 286  
|||||  
Db 338 TCTCCAAAGGGCAGCGCTTTCATCTCGTGTCTCTCCGTCACGACGAGTGTGTGAGG 279  
|||||  
QY 287 AGCTGGGCCCCATCTTACAGCTCATCTGTCAGATCAAGGGCAGCGTGGAGCATCCCCG 346  
|||||  
Db 278 AGCTGAGCCCCATCTTACAGCTCATCTGTCAGATCAAGGGCAGCGTGGAGCATCCCCA 219  
|||||  
QY 347 TGATCTCTGTGGGCAACAAGTCCGATGAGAGCGACGCGGAGGTGGACACGCCGAGGGCG 406  
|||||  
Db 218 TCATGCTGTGTGGGCAACAAGTGTGATGAGACGACGCGGAGGTGCACACCGCGAGGGCG 159  
|||||  
QY 407 AGCGCGTGGCCCCAAGAGGGAAGTGCCTTTTCATGAGACCTCGGCGCAAGATGAATACCA 466  
|||||  
Db 158 AGCGCGTGGCCCCAAGAGGGAAGTGCCTTTTCATGAGACCTCGGCGCAAGATGAATACCA 99  
|||||  
QY 467 ACGTCAAGAGCTCTTCCAGAGAGCTGTGACGCTGGAGACGCGCGGAACATGAGCTTCA 526  
|||||  
Db 98 ACGTGAAGAGCTGTTCAGAGAGCTGTGACGCTGGAGACGCGCGCGCATGAGCTTCA 39  
|||||  
QY 527 ACATGAGGCGCAAGCGCTCCGGGAGAGCAAGAGGAC 563  
|||||  
Db 38 GCGTGGACGGCAAGCGCTCCCAACAACATGTGGCGGC 2  
|||||

RESULT 7  
BE095936/c 456 bp mRNA linear EST 12-JUN-2000  
LOCUS  
DEFINITION UI-R-BU0-apd-d-07-0-UI.s1 UI-R-BU0 Rattus norvegicus cDNA clone  
BE095936  
ACCESSION UI-R-BU0-apd-d-07-0-UI 3', mRNA sequence.  
VERSION  
KEYWORDS BE095936.1 GI:8486867  
EST.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 456)  
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
TITLE Normalization and subtraction: two approaches to facilitate gene  
discovery  
JOURNAL Genome Res. 6 (9), 791-806 (1996)  
MEDLINE 97044477  
COMMENT Contact: Soares, MB  
Program for Rat Gene Discovery and Mapping  
University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: msouares@blue.weeg.uiowa.edu  
Oligo-dT track not found, Not I site shown in beginning of sequence  
is likely internal to the message. cDNA library Preparation: M.B.  
Soares Lab Clone distribution: clones will be available through  
Research Genetics (www.resgen.com)  
Seq primer: M13 Forward  
POLYA-No.

FEATURES  
Source  
Location/Qualifiers  
1..456  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-BU0-apd-d-07-0-UI"  
/clone\_1lb="UI-R-BU0"  
/dev\_stage="adult"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pUT3D-Pac (Pharmacia) with a modified  
polylinker. Site\_1: Not I; Site\_2: Eco RI; The UI-R-BU0  
library is a subtracted library derived from a mixture of  
eye and ganglia tissues. For a detailed description of the  
library from which this clone was derived, please visit  
our web site at [ratest.eng.uiowa.edu](http://ratest.eng.uiowa.edu). The subtraction has

been previously described in (Bonaldi, Lennon and Soares,  
Genome Research 6:791-806, 1996)  
TAG\_SEQ=None found"

Query Match 63.9%; Score 381.4; DB 10; Length 456;  
Best Local Similarity 89.9%; Pred. No. 5.8e-67;  
Matches 409; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

BASE COUNT 66 a 142 c 140 g 108 t

ORIGIN

109 ATCCCAACATTCAGGACCTTACCGGAGGTATCATGTCGACAGACCGTGTGACG 168  
|||||  
456 ATCCCAACATTCAGGACCTTACCGGAGGTATCATGTCGACAGACCGTGTGACG 397  
|||||  
169 CTGACATCATCAGACACACCGGACGACGACGACGACGACGACGACGACGACG 228  
|||||  
396 CTGACATCATCAGACACACCGGACGACGACGACGACGACGACGACGACGACG 337  
|||||  
229 TCCAAAGGCGCAGCTTCTCTGTTGTTCTCGTCCGTCACGACGACGACGACG 288  
|||||  
336 TCCAAAGGCGCAGCTTCTCTGTTGTTCTCGTCCGTCACGACGACGACGACG 277  
|||||  
289 CTGAGGCGCCTTCTCTGTTGTTCTCGTCCGTCACGACGACGACGACGACG 348  
|||||  
276 CTGAGGCGCCTTCTCTGTTGTTCTCGTCCGTCACGACGACGACGACGACG 217  
|||||  
349 ATGCTGTGGGCAACAGATGCGATGAGACGACGACGACGACGACGACGACG 408  
|||||  
216 ATGCTGTGGGCAACAGATGCGATGAGACGACGACGACGACGACGACGACG 157  
|||||  
409 GCGGTGGGCAACAGATGCGATGAGACGACGACGACGACGACGACGACGACG 468  
|||||  
156 GCGGTGGGCAACAGATGCGATGAGACGACGACGACGACGACGACGACGACG 97  
|||||  
469 GTCAGAGAGCTTCTCTGTTGTTCTCGTCCGTCACGACGACGACGACGACG 528  
|||||  
96 GTGAAGAGAGCTTCTCTGTTGTTCTCGTCCGTCACGACGACGACGACGACG 37  
|||||  
529 ATGACGCGCAACGCTTCTCTGTTGTTCTCGTCCGTCACGACGACGACGACG 563  
|||||  
36 GTGAGCGGCAACGCTTCTCTGTTGTTCTCGTCCGTCACGACGACGACGACG 2

RESULT 8  
A1764459/c 422 bp mRNA linear EST 25-JUN-1999  
LOCUS A1764459  
DEFINITION UI-R-Y0-abj-f-09-0-01.s1 UI-R-Y0 Rattus norvegicus cDNA clone  
ACCESSION A1764459  
VERSION A1764459.1 GI:5210352  
KEYWORDS EST.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 422)  
Bonaldi,M.F., Lennon,G. and Soares,M.B.  
Normalization and subtraction: two approaches to facilitate gene  
discovery  
Genome Res. 6 (9), 791-806 (1996)  
97044477  
Contact: Soares, MB  
Program for Rat Gene Discovery and Mapping  
University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: msoares@blue.weeg.uiowa.edu  
Oligo-qt track not found, Not 1 site shown in beginning of sequence  
is likely internal to the message. cDNA Library Preparation: M.B.  
Soares Lab Clone distribution: clones will be available through  
Research Genetics (www.resgen.com)

Seq primer: M13 Forward  
POLYA-No.  
FEATURES  
source

Location/Qualifiers  
1..422  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-Y0-abj-f-09-0-01"  
/clone\_lib="UI-R-Y0"  
/dev\_stage="adult"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker. Site\_1: Not I; Site\_2: Eco RI; The UI-R-Y0  
library is a subtracted library derived from an  
individually-tagged normalized whole-eye (minus the lens)  
library. The driver for the subtraction consisted of a  
pool of all previous libraries (UI-R-0, UI-R-A1, UI-R-E0,  
UI-R-E1, UI-R-C0, and UI-R-C1). The tag is a string of  
3-5 nucleotides present between the Not I site and the  
oligo-qt track which allows identification of the library  
of origin of a clone within the mixture. The subtracted  
library (UI-R-Y0) was constructed as follows: PCR  
amplified cDNA inserts from previous library clones from  
which 3' ESTs had been derived were used as a driver in a  
hybridization with the normalized whole-eye library in  
the form of single-stranded circles. The remaining  
single-stranded circles (subtracted library) was purified  
by hydroxyapatite column chromatography, converted to  
double-stranded circles and electroporated into DH10B  
bacteria (Life Technologies) to generate the UI-R-Y0  
library. This procedure has been previously described  
(Bonaldi, Lennon and Soares, Genome Research 6: 791-806,  
1996)

BASE COUNT 63 a 133 c 128 g 98 t

ORIGIN

Query Match 59.3%; Score 353.8; DB 9; Length 422;  
Best Local Similarity 90.0%; Pred. No. 2e-61;  
Matches 379; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

143 TCACGTCGCAACAGAGCGTGTGACGCTGCAATCATCAGACACACCGGACGACG 202  
|||||  
422 TCACGTCGCAACAGAGCGTGTGACGCTGCAATCATCAGACACACCGGACGACG 363  
|||||  
203 TCCCGGCGATGACGCGCTGTCATCTCCAAAGGCGGACGCTTCATCTGTTCTCG 262  
|||||  
362 TCCCGGCGATGACGCGCTGTCATCTCCAAAGGCGGACGCTTCATCTGTTCTCG 303  
|||||  
263 TCACGTCGCAACAGAGCGTGTGACGCTGCAATCATCAGACACACCGGACGACG 322  
|||||  
302 TCACGTCGCAACAGAGCGTGTGACGCTGCAATCATCAGACACACCGGACGACG 243  
|||||  
323 AGGCGAGCGTGTGACGCGCTGTCATCTCCAAAGGCGGACGCTTCATCTGTTCTCG 382  
|||||  
242 AGGCGAGCGTGTGACGCGCTGTCATCTCCAAAGGCGGACGCTTCATCTGTTCTCG 183  
|||||  
383 GGGAGGTGACAGCGCGGACGCGGCTGTCATCTCCAAAGGCGGACGCTTCATCTG 442  
|||||  
182 GGGAGGTGACAGCGCGGACGCGGCTGTCATCTCCAAAGGCGGACGCTTCATCTG 123  
|||||  
443 AGACCTCGGCGCAAGATGAATCAACGTCAGAGGCTTCCAGAGCTGTGACGCTGG 502  
|||||  
122 AGACCTCGGCGCAAGATGAATCAACGTCAGAGGCTTCCAGAGCTGTGACGCTGG 63  
|||||  
503 AGACGCGCGGCAAGATGAATCAACGTCAGAGGCTTCCAGAGCTGTGACGCTGG 562  
|||||  
62 AGACGCGCGGCAAGATGAATCAACGTCAGAGGCTTCCAGAGCTGTGACGCTGG 3  
|||||  
563 C 563

Db 2 C 2

RESULT 9  
AM533412/c 419 bp mRNA linear EST 06-MAR-2000  
LOCUS  
DEFINITION  
UI-R-BU0-amb-f-04-0-UI.s1 UI-R-BU0 Rattus norvegicus CDNA clone  
UI-R-BU0-amb-f-04-0-UI 3', mRNA sequence.  
AM533412  
ACCESSION  
AM533412.1 GI:7175826  
KEYWORDS  
EST.  
SOURCE  
ORGANISM  
Norway rat.  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 419)  
REFERENCE  
AUTHORS  
TITLE  
Normalization and subtraction: two approaches to facilitate gene discovery  
Genome Res. 6 (9), 791-806 (1996)  
JOURNAL  
MEDLINE  
COMMENT  
Contact: Soares, MB  
Program for Rat Gene Discovery and Mapping  
University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: msoares@blue.weeg.uiowa.edu  
Oligo-dT track not found, Not 1 site shown in beginning of sequence  
is likely internal to the message. CDNA Library Preparation: M.B.  
Soares Lab Clone distribution: clones will be available through  
Research Genetics (www.resgen.com)  
Seq primer: M13 Forward  
POLYA-No.

FEATURES  
source  
Location/Qualifiers  
1..419  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-BU0-amb-f-04-0-UI"  
/clone\_lib="UI-R-BU0"  
/dev\_stage="adult"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker. Site 1: Not 1; Site 2: Eco RI; The UI-R-BU0  
library is a subtracted library derived from a mixture of  
eye and ganglia tissues. For a detailed description of the  
library from which this clone was derived, please visit  
our web site at [ratest.eng.uiowa.edu](http://ratest.eng.uiowa.edu). The subtraction has  
been previously described in (Bonaldo, Lennon and Soares,  
Genome Research 6:791-806, 1996)  
TAG\_LIB="UI-R-BU0"  
TAG\_TISSUE="ganglia"  
TAG\_SEQ="GCAGAC"

BASE COUNT 62 a 132 c 128 g 97 t

ORIGIN

Query Match 58.5%; Score 349.2; DB 10; Length 419;  
Best Local Similarity 89.7%; Pred. No. 1.7e-60;  
Matches 375; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

Db 146 GCTGCGACAAGAGGCTGTCAGAGTCACAGACACGCGGACCGACAGTCC 205  
|||||  
Db 419 GCTGTGACAAGAGGCTGTCAGACTCCAGATCAGGACACGCGGACCGACAGTCC 360  
|||||  
QY 206 CGGCGATGACAGCGCTGTCATCTCCAGGCGACGCTTCATCTGCTTCCTCA 265  
|||||  
Db 359 CAGCATGACAGCGGCTGTCATCTCCAGGCGACGCTTCATCTGCTTCCTCA 300  
|||||  
QY 266 CCAGCAGACAGTCTGTCAGAGAGCTGGGCGCATCTACAGCTCATGTCAGATCAAG 325  
|||||  
Db 299 CCAGCAGACAGTCTGTCAGAGAGCTGAGCCCATCTACAGCTCATGTCAGATCAAG 240  
|||||

QY 326 GCAGCGTGAGAGACATCCCGTGATGCTGTGGGCAACAAGTGCATGAGACGACGGG 385  
|||||  
Db 239 GCAGCGTGAGAGACATCCCGTGATGCTGTGGGCAACAAGTGCATGAGACGACGGG 180  
|||||  
QY 386 AGGTGACAGCGCGGAGCGGACGCGGTGGCCCAAGAGTGAAGTGCCTTCATGAGAGA 445  
|||||  
Db 179 AGGTGACAGCGCGGAGCGGACGCGGTGGCCCAAGAGTGAAGTGCCTTCATGAGAGA 120  
|||||  
QY 446 CCTGCGCAAGATGAACTCAACAGCTTCAGAGAGCTTCAGAGAGCTGTCAGAGAGA 505  
|||||  
Db 119 CCTGCGCAAGATGAACTCAACAGCTTCAGAGAGCTTCAGAGAGCTGTCAGAGAGA 60  
|||||  
QY 506 CGCGCGGAAATGAGAGCTTCACATGACGAGCAAGCGCTCCGGAAGAGAGAGAG 563  
|||||  
Db 59 CGCGCGGAGAGCTGACGCTTCAGAGAGCTGAGCGCAAGCGCTCCACAAACATGCGCGCC 2  
|||||

RESULT 10  
BG702913 801 bp mRNA linear EST 07-MAY-2001  
LOCUS  
DEFINITION  
602684858F1 NIH\_MGC\_95 Homo sapiens CDNA clone IMAGE:4817247 5',  
mRNA sequence.  
BG702913  
ACCESSION  
BG702913  
VERSION  
BG702913.1 GI:13974731  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 801)  
REFERENCE  
AUTHORS  
TITLE  
NIH-MGC <http://mgc.ncl.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
JOURNAL  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: Miklos Palokovits, M.D., Ph.D.  
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki  
Toshiyuki and Piero Carninci (RIKEN)  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LRAM10717 row: d column: 16  
High quality sequence stop: 760.

FEATURES  
source  
Location/Qualifiers  
1..801  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4817247"  
/clone\_lib="NIH\_MGC\_95"  
/lssue\_type="hippocampus"  
/lab\_host="DH10B"  
/note="Organ: brain; Vector: pBluescriptR (modified  
pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtagag  
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTNN-3',  
size selected for average insert size 2.5 kb and  
normalized to R0T 5. This is a primary library enriched  
for full-length clones and constructed using the  
Cap-trapper method (Carninci, in preparation). Library  
constructed by M. Brownstein (NIH/NHGRI, National  
Institutes of Health). Note: this is a NIH\_MGC Library."

BASE COUNT 184 a 223 c 252 g 142 t

ORIGIN

Query Match 56.9%; Score 339.4; DB 12; Length 801;  
Best Local Similarity 78.3%; Pred. No. 1.7e-58;  
Matches 458; Conservative 0; Mismatches 121; Indels 6; Gaps 4;

Db 1 ATGCCGAGACAGACTAACGATTACCGCTGTGCTGGGGCGGCGCGTGGGCAAG 60  
|||||  
Db 152 ATGCTTGACAGAGAGTAACGATTACCGGCTGCTTTGGGGCTGGCGGTGGCAAG 211  
|||||







Toshiyuki and Piero Carninci (RIKEN)  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNLN at:  
<http://image.llnl.gov>  
Plate: LLM10703 row: h column: 11  
High quality sequence stop: 701.

FEATURES  
source  
1..709  
Location/Qualifiers

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4811962"  
/clone\_id="NIH\_MGC\_95"  
/tissue\_type="hippocampus"  
/lab\_host="DH10B"  
/note="Organ: brain; Vector: pBluescriptR (modified  
pBluescript KS+); Site\_1: BamHI; Site\_2: SalI-XhoI (gtcgag  
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',  
size-selected for average insert size 2.5 kb and  
normalized to ROT 5. This is a primary library enriched  
for full-length clones and constructed using the  
Cap-trapper method (Carninci, in preparation). Library  
constructed by M. Brownstein (NIH/NHGRI, National  
Institutes of Health). Note: this is a NIH\_MGC Library."

BASE COUNT 168 a 191 c 226 g 124 t

ORIGIN

Query Match 52.1%; Score 311.2; DB 12; Length 709;  
Best Local Similarity 79.5%; Pred. No. 7.9e-53;  
Matches 431; Conservative 0; Mismatches 103; Indels 8; Gaps 5;

OY 1 ATGCCGGAGACAGATTACCGGCTGGTTCGGGGCGGGCGGGGCAAG 60  
DB 151 ATGCTTGAAGCAGATTACCGGCTGGTTCGGGGCGGGGCAAG 209  
OY 61 AGCTCGCTGTGCTCGCTTCGTAAGGGCAGTTCGCGACACCTACATCCACCATC 120  
DB 210 AGCTCCCTGGTGTGAGGTT-GTGAAGCACAATTCCGGGAGAGGTACATCCGACGGTG 268  
OY 121 GAGACACCTTACCGGAGGTATGACTGCGACAAGAGCGTGTGCGAGCTGCAGATCACA 180  
DB 269 GAAGACACCTACCGGAGGTATGACTGCGACAAGAGCATATGCACATTCAGATCAC 328  
OY 181 GACACCAACCGGAGCAGCATTCGCGCATGCGCGCTGCATCTCCAAAGGGCCAC 240  
DB 329 GACACGAGGGGAGCCACCATTCGCGCATGCGCGCTGCATCTCCAAAGGGCCAC 388  
OY 241 GCCTTCATCTGTGTTCTCCGTCACAGACAGTGTGAGAGCTGGGGCCATC 300  
DB 389 GCCTTCATCTGTGTTCTCCGTCACAGACAGTGTGAGAGCTGGGGCCATC 448  
OY 301 TACAAGCTCATCGTCAGATCAAGGGCAGCTGAGACATCCCGTATGCTGTGGGC 360  
DB 449 TACGACACAAATCTCGAGATCAAGGGGAGCGTGGAGACATCCCATCATGCTGTGGGC 508  
OY 361 AACAGTCCGATGAGA---CGCAGCGGAGGTGAGACCGCGAGGGCAGCGGTGGCC 417  
DB 509 AACAGTGTGATGAGAGCCGCGAGCGGAGGTGAGACCGCGAGGGCAGCGGTGGCC 566  
OY 418 CAAGAGTGAAGTGCCTTTATGAGAGCTCGGCCAGATGAATACAGCTCAAGAG 477  
DB 567 CGCAGATGGAAGTGTGCTTATGAGAGCTCGGCCAGATGAATACAGCTCAAGAG 626  
OY 478 CTCTTCCAGAGCTGTGAGAGCGCGCGGAACATGAGCCTCAACATGAGCGGC 537  
DB 627 CTTTTCAGAGAGCTGTCAAACTGGAGAGCGCAGG-ACGTGAGTCTCCAGATCGACGG 685  
OY 538 AA 539  
DB 686 AA 687

Search completed: March 23, 2003, 17:12:31  
Job time : 3887.38 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 23, 2003, 15:14:42 ; Search time 157.566 Seconds  
(Without alignments)  
2938.192 Million cell updates/sec

Title: US-09-873-546-4

Perfect score: 597

Sequence: 1 atgcgcgaacagagtaacga.....gcaatgcacccatcgtga 597

Scoring table: IDENTITY NUC

Searched: 538826 seqs, 387737923 residues

Total number of hits satisfying chosen parameters: 1077652

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: /cgnt2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgnt2\_6/ptodata/2/pubpna/PCY\_NEM\_PUB.seq:\*  
3: /cgnt2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
4: /cgnt2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgnt2\_6/ptodata/2/pubpna/US07\_NEM\_PUB.seq:\*  
6: /cgnt2\_6/ptodata/2/pubpna/PCYUS\_PUBCOMB.seq:\*  
7: /cgnt2\_6/ptodata/2/pubpna/US08\_NEM\_PUB.seq:\*  
8: /cgnt2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgnt2\_6/ptodata/2/pubpna/US09\_NEM\_PUB.seq:\*  
10: /cgnt2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgnt2\_6/ptodata/2/pubpna/US10\_NEM\_PUB.seq:\*  
12: /cgnt2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgnt2\_6/ptodata/2/pubpna/US60\_NEM\_PUB.seq:\*  
14: /cgnt2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	234.4	39.3	452	10	US-09-960-352-4253
2	234.4	34.3	454	10	US-09-960-352-5830
3	151	25.3	3082	10	US-09-778-963A-1
4	137.8	23.1	551	10	US-09-765-298A-25
5	137.8	23.1	570	9	US-09-359-585-7
6	137.8	23.1	570	10	US-09-765-298A-27
7	136.2	22.8	570	12	US-10-104-484-1
8	136.2	22.8	570	12	US-10-104-484-3
9	118.2	19.8	3346	9	US-09-764-868-67
10	118	19.8	688	9	US-09-764-868-490
11	107.8	18.1	1119	9	US-10-067-813-1
12	107.8	17.2	1085	9	US-10-067-813-3
13	100.4	16.6	11221	10	US-09-778-963A-3
14	99	16.6	405	10	US-09-960-352-10273
15	97	16.2	368	10	US-09-864-761-21643
16	84.2	14.1	612	10	US-09-972-529-3
17	84.2	14.1	1224	10	US-09-972-529-1
18	84.2	14.1	3192	10	US-09-788-654A-1
19	81.2	13.6	3936	10	US-09-919-172-49

20	80.8	13.5	752	9	US-09-764-868-501	Sequence 501, App
21	80.4	13.5	1271	12	US-10-044-090-110	Sequence 110, App
22	72.4	12.1	925	10	US-09-967-736-4	Sequence 4, App1
23	72	12.1	969	10	US-09-801-368-287	Sequence 287, App
24	67	11.2	407	10	US-09-960-352-11732	Sequence 11732, A
25	67	11.2	267	10	US-09-925-300-678	Sequence 678, App
26	66.4	11.1	942	10	US-09-962-832-256	Sequence 256, App
27	66.2	11.1	2021	9	US-09-764-868-8	Sequence 88, App1
28	66.2	11.0	3257	10	US-09-817-198A-1	Sequence 1, App1
29	65.6	11.0	4083	10	US-09-817-182-1	Sequence 1, App1
30	65	10.9	952	10	US-09-917-800A-1389	Sequence 1389, App
31	64	10.7	930	10	US-09-801-368-285	Sequence 285, App
32	63.8	10.7	2183	10	US-09-917-800A-1104	Sequence 1704, App
33	63.6	10.7	1191	12	US-10-044-090-111	Sequence 111, App
34	61.2	10.3	2291	9	US-10-098-841-237	Sequence 237, App
35	61.2	10.3	2418	9	US-09-764-868-84	Sequence 84, App1
36	61.2	10.3	2479	9	US-09-764-868-69	Sequence 69, App1
37	60.8	10.2	1157	9	US-10-108-605-44	Sequence 44, App1
38	60.6	10.2	1443	10	US-09-962-832-224	Sequence 224, App
39	60.2	10.1	566	9	US-09-764-868-507	Sequence 507, App
40	59.6	10.0	4017	10	US-09-880-107-3029	Sequence 3029, App
41	57.2	9.6	1007	10	US-09-822-849A-563	Sequence 563, App
42	57	9.5	1274	10	US-09-925-302-91	Sequence 91, App1
43	56.4	9.4	385	10	US-09-960-352-13143	Sequence 13143, A
44	56.4	9.4	1077	9	US-09-764-868-497	Sequence 497, App
45	56.4	9.4	1364	9	US-09-764-868-77	Sequence 77, App1

#### ALIGNMENTS

RESULT 1  
US-09-960-352-4253  
Sequence 4253, Application US/09960352  
Patent No. US20020137139A1  
GENERAL INFORMATION:  
APPLICANT: Warren, Wesley C.  
APPLICANT: Tao, Nengping  
APPLICANT: Bvatt, John C.  
APPLICANT: Mathiaslagu, Nagappan  
TITLE OF INVENTION: MUSCLE ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
FILE REFERENCE: 16511.006/37-21(10298)C  
CURRENT APPLICATION NUMBER: US/09/960.352  
CURRENT FILING DATE: 2001-09-24  
NUMBER OF SEQ ID NOS: 15112  
SEQ ID NO 4253  
LENGTH: 452  
TYPE: DNA  
ORGANISM: Bos taurus  
OTHER INFORMATION: Clone ID: 18-LIB34-086-Q1-E1-E5  
US-09-960-352-4253

Query Match 39.3%; Score 234.4; DB 10; Length 452;  
Best Local Similarity 71.5%; Pred. No. 2.1e-43;  
Matches 323; Conservative 0; Mismatches 126; Indels 3; Gaps 1;  
OY 34 GTGTTGGGGGGGGGGGGGGGCAAGAGCTCGTGGTCTGCTGCTGTAAGGCAAG 93  
DB 1 GTGATCGGCTCGGGGGGGGGGCAAGAGCGCTGTTGTTGCAAGGTGGCGGCAAG 60  
OY 94 TTCCGCGACCTTACATATCCCAATGAGAGACCTTCCGAGAGTATAGCTGCAGC 153  
DB 61 TTCCGAGGCGGTACTTCCGACCATCAATGATCTTACGCGAGGCGCTAGCTGCAGC 120  
OY 154 AAGAGCGTGTGACGCTGCAGATCAAGACACACCGGAGCGACCATTTCCCGCATG 213  
DB 121 CACAGGCGGGGTCCTCATCATCAACACACCGGTCGCGCGCTACCGGGGCTG 180  
OY 214 CAGCGCTTGCATCTTCCAAAGGCGCAGCGCTTATCTTGGTGTTCCTGATACGCAAG 273  
DB 181 CAGCGCTTGCATCTTCCAAAGGCGTACGCGCTTATCTTGGTGTTCCTGATACGCAAG 240







QY 379 CAGCGGAGGTGACACGCCGCGAGCGGTGGCCCAAGAGTGAAGTGCCTTTC 438  
DB 364 GCACGACGTGTGGATATCTGGCAGGCTCAGACCTGCCGGAAGCTACGCCATCCCCCTAC 423  
QY 439 ATGAGAGCTCGGCCAAGATGAACACTCAACAGTCAAGAGCTCTTCCAGAGCTGCTG 495  
DB 424 ATCGAGACCTCGGCCAAGACCGGCGAGGAGTGGAGAGTCCCTTTCACAGCTTGGTG 480

## RESULT 9

US-09-764-868-67  
; Sequence 67, Application US/09764868  
; Patent No. US2002016871A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT32  
; CURRENT APPLICATION NUMBER: US/09/764,868  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - refer to PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1510  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO: 67  
; LENGTH: 3346  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (2787)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: SITE  
; LOCATION: (2795)  
; OTHER INFORMATION: n equals a,t,g, or c  
US-09-764-868-67

Query Match 19.8%; Score 118.2; DB 9; Length 3346;  
Best Local Similarity 52.5%; Pred. No. 1.6e-17;

Matches 258; Conservative 0; Mismatches 233; Indels 0; Gaps 0;

QY 9 ACAGAGTAACGATTAACCGCGTGTGTTCGGGGGCGGCGGTGGCCAAAGCTCGCT 68  
DB 188 ACTCATGAGGAGATACAAAGTAGTGTGTAGGAGTGGAGGGGTGGCAAAATCTCCCT 247  
QY 69 GGTGTGCGCTTCGTGAGGAGCGTTCGCGCACCTACATCCACATGAGGAGAC 128  
DB 248 TACTGTGACGTTTGTACTGAGGACTTTCATTTGAAATATGACCCCAATTTGACACCGC 307  
QY 129 CTACCGGAGGTGATCAGCTGCGACAAGAGCTGTGACAGCTGCAGATCACAGACAC 188  
DB 308 CTACCGCAAGAGATGAGAGTGAATCTTCCCTCCGCTGAGAAATTTCTGACACCGC 367  
QY 189 CGGACGCCACAGTTCGCGGCATGACGCGCTGTCCATCTCCAGAGGCCAGCTTCAT 248  
DB 368 AGGAAGTGAAGCTTTCCTCATGAGAGATCTCTACATCAAAAAGCGCCAAAGTTTCAT 427  
QY 249 CCGGTGTCTCCGTCACACGACGAGCGTGGAGAGAGCTGGGGCCATCTACAGCT 308  
DB 428 CCGTGTGTATAGCTGTATATCAACAGTCTTTTCAGAGATCAACCCCAATTTGAGAGTCA 487  
QY 309 CATCTGTCAGATCAAGGAGCGGTGGAGACATCCCGTATGCTCTGTGGCAACAAGT 368  
DB 488 AATTGTGACAGTGAAGAGATGATAAAGTCCCATTAATCTGTAGAAATTAAGTGA 547  
QY 369 CGATGAGACGACGCGGAGGTGACACGCGGAGGCGGTGGCCCAAGATGGAA 428  
DB 548 TCTGGAACCAAGAAAGAGTATGTCTTCAGAAAGCAGAGCTCTGCTCAAAAGTGGG 607  
QY 429 GTGCGCTTTCATGAGACCTCGGCCAAGATGAACATAAGTCAAGAGCTCTTCAGGA 488  
DB 608 CTGTCTTTTATGAGACATCGGCAAAAGTAAATCAATGTGATGAATTTTCTCTGA 667  
QY 489 GCTGTGACGC 499

DB 668 GATCTGTACAGC 678

## RESULT 10

US-09-764-868-490  
; Sequence 490, Application US/09764868  
; Patent No. US2002016871A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT32  
; CURRENT APPLICATION NUMBER: US/09/764,868  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - refer to PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1510  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO: 490  
; LENGTH: 688  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (579)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: SITE  
; LOCATION: (610)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: SITE  
; LOCATION: (669)  
; OTHER INFORMATION: n equals a,t,g, or c  
US-09-764-868-490

Query Match 19.8%; Score 118; DB 9; Length 688;  
Best Local Similarity 52.1%; Pred. No. 1.4e-17;

Matches 256; Conservative 1; Mismatches 234; Indels 0; Gaps 0;

QY 9 ACAGAGTAACGATTAACCGCGTGTGTTCGGGGGCGGCGGTGGCCAAAGAGCTCGCT 68  
DB 181 ACTCATGAGGAGATACAAAGTAGTGTGTAGGAGTGGAGGGGTGGCAAAATCTCCCT 240  
QY 69 GGTGTGCGCTTCGTGAGGAGCGTTCGCGCACCTACATCCCAATGAGGAGAC 128  
DB 241 TACTGTGACGTTTGTACTGAGGACTTTCATTTGAAATATGACCCCAATTTGACATTT 300  
QY 129 CTACCGGAGGTGATCAGCTGCGACAAGAGCTGTGACAGCTGCAGATCACAGACAC 188  
DB 301 CTACCGCAAGAGATGAGAGTGAATCTTCCCTCCGCTGAGAAATCTTGGACACCGC 360  
QY 189 CGGACGCCACAGTTCGCGGCATGACGCGCTGTCCATCTCCAGAGGCCAGCTTCAT 248  
DB 361 AGGAAGTGAAGCTTTCCTCATGAGAGATCTCTACATCAAAAAGCGCCAAAGTTTCAT 420  
QY 249 CCGGTGTCTCCGTCACACGACGAGCTGTGAGAGAGCTGGGGCCATCTACAGCT 308  
DB 421 CCGTGTGTATAGCTGTATATCAACAGTCTTTTCAGAGATCAAGCAATTTGAGAGATCA 480  
QY 309 CATCTGTCAGATCAAGGAGCGGTGGAGACATCCCGTATGCTCTGTGGCAACAAGT 368  
DB 481 AATTGTGACAGTGAAGAGATGATAAAGTCCCATTAATCTCKAGTGAAGAAATTAAGTGA 540  
QY 369 CGATGAGACGACGCGGAGGTGGAGACAGCGCGAGGCGCAGGCGTGGCCCAAGATGGAA 428  
DB 541 TCTGGAACCAAGAAAGAGTATGTCTTTCAGAAAGCAGAAAGTCTGTGCTCAAAATGGG 600  
QY 429 GTGCGCTTTCATGAGACCTCGGCCAAGATGAACATAAGTCAAGAGCTCTTCAGGA 488  
DB 601 CTGTCTTTTATGAGACATCGGCAAAAGTAAATCAATGTGATGAATTTTCTCTGA 660  
QY 489 GCTGTGACGC 499  
DB 661 GATCTGTACAGC 671

```

? SEQ ID NO 1
? LENGTH: 1119
? TYPE: DNA
? ORGANISM: Murinae gen. sp.
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (19)..(642)
?-S-10-067-813-1
```

QY	29	TGGGGTGTTCGGGGCGGGCGGCTGGGGCAAGAGCTCGCTGTGTCCTCGCTTCGTGAAG	88
Db	65	TAGTAGTGGTGGGAGATGTGTGTGGGCAAGAGTGGCGCTCAATTCAGTTTTTCCAGA	124
QY	89	GCAGTTCGCGGACACCTACATCCCAACCATCGAGAGACCTTACCGCAGGTATCAGCT	148
Db	125	AGATCTTTGTGCTGAGCTACAGACCCCAACATGAAAGCTCTTACCTGAACATACAGAGA	184
QY	149	GGCAGCAAGACGTTGTGCACGCTGCAGATCAACAGACACACCGCAGCAGACAGTCCGG	208
Db	185	TTGCACATCAGTGGGCGCATCTTGGATGTTCGACACAGCGCGGCAAGAGAGTTCACTG	244
QY	209	CCATGTCAGCGCCCTGTGCTATCTCCAGGGGCCACGCTTCACTCTGTGTTCCTCGTACCA	268
Db	245	CCATGCGGGAACTATACATATCGCAACAGGGATGGTTCCTATTTGTTACTCTCGTACCG	304
QY	269	GCAAGCAGTGGCTGGAGAGAGCTGGGGCCACTTACAAGCTCATCTGTCAGCATCAAGGCA	328
Db	305	ACAAGGCAAGCTTCTGAGACAGCTGGACCGGCTTCACCAGCACTTTCTGCGTGAAGSACA	364
QY	329	GCGTGGAGGACATCCCGCTGATGCTGTTGGGCAACAAGTCCGATGAGACGCAAGC---GGG	385
Db	365	G---GGGATCATTTCCCAATGATCTCTGTGGCCAAACAAGGTGATCGATGCACTTAAGGA	421
QY	386	AGGTGACACGCGCAGAGGCGCAGCGGTGGCCCAAGAGTGGAAAGTCCGGCTTCATAGAGA	445
Db	422	AAGTCAACAGAGGACCAAGAAAAAGAAATGCAACCAATACATATATCCATATATAGAGA	481
QY	446	CCCTGGGCAAG 456	
Db	482	CCAGTGGCAAG 492	

RESULT 12  
US-10-067-813-3  
Sequence 3, Application US/10067813  
Patent No. US20020156013A1  
GENERAL INFORMATION:  
APPLICANT: Remauld, Jean-Christophe  
APPLICANT: Louahed, Jamila

Query Match	17.2%;	Score 102.6;	DB 9;	Length 1085;
Best Local Similarity	56.2%;	Pred. No. 3.9e-14;		
Matches 214; Conservative	0;	Mismatches 164;	Indels 3;	Gaps 1

QY	22	TACGCGCTGGTGGTGGTTCGGGGGGGGGGGGGGTGGGCAAGAGAGCTGGGTGGTGGCTTC	81
Db	141	TACAAAGCTGGTGGTGGTGGGGGATGGGGGATGGGGCAAAAGTGGCCCTACCAATCCAGTTT	200
QY	82	GTGAAGGGCACGTTCCGGGACACCTTACATCCCACTATCGAGACACTTACCGGACAGTG	144
Db	201	TTCCAGAAAGATCTTTGGTCCGACTATGAGCCCAACCATTTGAAGACTCTTACTGAAACAT	260
QY	142	ATCAGCTGGCACAAGAGCGTGTGCAGCGCTGCAGATCACAGACACCAACCGGACAGCACAG	201
Db	261	ACGGAGATTGCACATCAATGAGGCCATCTTGAGACTTCTGCAACAGCTGGGACAGAGAA	320
QY	202	TTCCGCGGACATGACGCGCCCTGTCCATCTCCAAAGGGCACGCGCTTCACTCGTGGTTCTCC	261
Db	321	TTTACGCCCCATGCGGAGCAATACATCGGACGGGGGATGGCTTCTCATCGTCTACTCC	380
QY	262	GTCAACGACGAAGCAGTGGCTGTGGAGAGACTGGGGGCCATTTACAAGCTCATCGTGCAGATC	321
Db	381	GTCACTATCAAGGCGACACTTTGTAGACAGCTGTGACCGCTTCCACCAAGCTTATCTTGGCGTTC	440
QY	322	AAGGGCAGCGTGAAGAGACATCCCGTGATCTCTGTGGCAACAAAGTGCATGTAGAGACAG	381
Db	441	AAAGACAG--GGAGTCATCTCCGATGATCTCTGTGGCAACAAAGGTGCATTTGATGACAC	497
QY	382	CGGAGGTGACACAGCGCGAG	402
Db	498	TTGAGGAAGATCACCAAGAG	518

```

, RESULT 13
, US-09-778-963A-3
, Sequence 3, Application US/09778963A
, Patent No. US200201172A1
, GENERAL INFORMATION:
, APPLICANT: MEELAM, Beena et al
, TITLE OF INVENTION: ISOLATED HUMAN RAS-LIKE PROTEINS,
, TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING THESE HUMAN RAS-LIKE
, FILE REFERENCE: C1001112
, CURRENT APPLICATION NUMBER: US/09/778, 963A
, CURRENT FILING DATE: 2001-02-08
, NUMBER OF SEQ ID NOS: 5
, SOFTWARE: FastSeq for Windows Version 4.0
, SEQ ID NO 3
, LENGTH: 11221
, TYPE: DNA
, ORGANISM: Homo sapiens
, RS-09-778-963A-3

```

Query Match	16.8%	Score 100.4	DB 10	Length 11221
Best Local Similarity	62.2%	Pred. No. 1.8e-13		
Matches 158	Conservative 0	Mismatches 96	Indels 0	Gaps 0
QY	16 AACGATACCCGCGTGGTGTGTTGGGGCGGGCGGGGCGAGACGCTGCSTGTGCTG 75			
Db	3051 AACCTCATCCCGCATGGTGTGCTGTGGCTCCCTGTGGGGGCGCAAGCTTCATCTGTGCT 3110			
QY	76 CGCTTCGTGAAGGGGACACTTCCGGCAGCACTACATCCCGACCATCGAGACACTTACCGG 135			
Db	3111 CGCTTCCCTCATGGCGCGCTTTGAGAGACCATGACACACCACCATCGAGAGACTTCCACCGT 3170			
QY	136 CAGGTGATCAGCTCGGAGCAAGAGCGGTGCAAGCTGCAGATCACAGACACCGACGCGAC 195			
Db	3171 AAGGATATACACATCCGGCGGAGCATGTACCAGCTGCAGATCTGTGATACCTGTGGCAAC 3230			
QY	196 CACCAAGTTCGGGCGCATCGACGCGCTGTCCATCTCCAAAGGCGCAGCCTTATCTCTGCTG 255			
Db	3231 CACCCCTTCCCGCGCATCGCAGGCTGTCCATCTCCACAGGTGAGGGGCCCACTGGTGCCTG 3290			
QY	256 TTTCCTGTCACCAAG 269			
Db	3291 GGCTGGGCGGCGAG 3304			

```

RESULT 14
US-09-960-352-10273
; Sequence 10273, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengping C.
; APPLICANT: Byate, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960.352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 10273
; LENGTH: 405
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (3), (23), (31)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: 44-LIB3058-050-Q1-K1-C8
-09-960-352-10273

```

Query Match	16.6%	Score 99	DB 10	Length 405
Best Local Similarity	63.0%	Pred. No. 2.1e-13		
Matches 153	Conservative 0	Mismatches 90	Indels 0	Gaps 0
QY	16 AACGATTACCGGCGTGTGTTCCGGGGCGGGCGTGGGCAAGACGTACGCTGTGTCTG	75		
Db	138 AACCTCTACCGCATGGTGTGCTCTCGGCTTCGTCCAAAGGTGGGCAAGCGCATTGCTG	197		
QY	76 GCCTTCGTGAAGGAGCGCTCCGGCAGACACTTCATCTCCACCATGGAGACACTACCGG	135		
Db	198 GCCTTCGTGACGGCGCGCTACGAGGACGCTTACACGCCCCACCATCGAACAAGCTTCCACCGC	257		
QY	136 CAGGTGATCAGCTGGCAGCAAGAGCGCTGTGCACAGCTGCACAGATCACAGACACCCAGCAGC	195		
Db	258 AAGTCTACTGTCAATTCGAGGCGAGAGGATCCAGCTGGACATCTCTCGACACGTACGGCAGAC	317		
QY	196 CACCAAGTTCGGCGGCATGTAGAGCGCTGTGTCAATCTCCAAAGGCGCAGGCTTACCTCGGTG	255		
Db	318 CGCGCATTCCTCCGACATGTGAGCGCCTGTGCATCTTACCGGAGAGCGTGTTCATCTTAGGG	377		
QY	256 TTC 258			

```

Db      378 TTC 380      111
RESULT 15
US-09-864-761-21643
: Sequence 21643, Application US/09864761
: Patent No. US20020048763A1
GENERAL INFORMATION:
: APPLICANT: Penn, Sharron G.
: APPLICANT: Rank, David R.
: APPLICANT: Hanzel, David K.
: APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aeomica X-1
CURRENT FILING DATE: US/09/864,761
PRIOR APPLICATION NUMBER: 2001-05-23
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,667
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 21643
LENGTH: 368
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AI022334.1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.8
OTHER INFORMATION: EXPRESSED IN LONG, SIGNAL = 3.2
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.6
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.9
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.9
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.7
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3
OTHER INFORMATION: NT HIT: q111418034, EVALUE 0.00e+00

```

OTHER INFORMATION: SWISSPROT HIT: P13856, EVALU 8.00e-12  
OTHER INFORMATION: EST\_HUMAN HIT: BE38944.1, EVALU 0.00e+00  
US-09-864-761-21643

Query Match 16.2%; Score 97; DB 10; Length 368;  
Best Local Similarity 65.4%; Pred. No. 5.7e-13;  
Matches 142; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 16 AACGATTACCGCGTGGTGTGTCGGGGCGGCGCGCTGGGCAAGAGCTGCTGTGCTG 75  
||| ||||| ||||| ||| ||| ||||| ||||| |||  
DB 149 AACTCATACCGCATGGTGGTGGTGGCTCTCGGGTGGCAAGAGCTCATCTGTCT 208  
||| ||||| ||||| ||| ||| ||||| ||||| |||  
QY 76 CGCTTGTGAAGGGCAAGTTCGCGACACCTACATCCCAACATCGAGACACCTACCG 135  
||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 209 CGCTTCTCAATGGCCGCTTTGAGACCAATACACCCACCATCGAGGACTTCACCGT 268  
||||| ||||| ||||| ||||| ||||| ||||| |||||  
QY 136 CAGGTGATCAGTGCAGACAAGAGCTGTGCACGCTGCAGATCAGAGACACACCGGACG 195  
||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 269 AAGGTATACACATCCGCGCGGCGCATGTACACAGCTGCACATCTCTGATACCTTGCGAAC 328  
||||| ||||| ||||| ||||| ||||| ||||| |||||  
196 CACCAAGTCCCGGCGCATGCGAGCGCGCTGTCCATCTCA 232  
||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 329 CACCCCTTCGCGCATGCGAGGCTGTCCATCTCA 365  
||||| ||||| ||||| ||||| ||||| ||||| |||||

Search completed: March 23, 2003, 17:17:10  
Job time : 166.566 secs



GenCore version 5.1.4 p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: March 23, 2003, 13:56:56 ; Search time 106.909 Seconds  
(without alignments)  
8438.824 Million cell updates/sec

Title: US-09-873-546-3

Perfect score: 31

Sequence: 1 gcggaattccatcagtgaggtcattgcc 31

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapept 1.0

Searched: 2054640 seqs, 1451402878 residues

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database :

GenBml: \*  
1: gb\_ba: \*  
2: gb\_htg: \*  
3: gb\_in: \*  
4: gb\_cm: \*  
5: gb\_ov: \*  
6: gb\_ph: \*  
7: gb\_pl: \*  
8: gb\_pr: \*  
9: gb\_ro: \*  
10: gb\_sts: \*  
11: gb\_sy: \*  
12: gb\_un: \*  
13: gb\_vl: \*  
14: gb\_ba: \*  
15: em\_fun: \*  
16: em\_hum: \*  
17: em\_in: \*  
18: em\_mu: \*  
19: em\_or: \*  
20: em\_ov: \*  
21: em\_ph: \*  
22: em\_pl: \*  
23: em\_ro: \*  
24: em\_sts: \*  
25: em\_un: \*  
26: em\_vl: \*  
27: em\_cm: \*  
28: em\_hum: \*  
29: em\_in: \*  
30: em\_mu: \*  
31: em\_or: \*  
32: em\_ov: \*  
33: em\_ph: \*  
34: em\_pl: \*  
35: em\_ro: \*  
36: em\_sts: \*  
37: em\_un: \*  
38: em\_vl: \*  
39: em\_cm: \*  
40: em\_hum: \*  
41: em\_in: \*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	23	74.2	774	9 AB076888	AB076888 Homo sapi
C 2	23	74.2	1249	6 AX430295	AX430295 Sequence
C 3	23	74.2	2827	9 AK096600	AK096600 Homo sapi
C 4	23	74.2	3061	6 AX430418	AX430418 Sequence
C 5	23	74.2	3391	9 BC030660	BC030660 Homo sapi
C 6	23	74.2	177540	9 AC006538	AC006538 Homo sapi
C 7	23	74.2	194456	2 AC113495	AC113495 Mus muscu
C 8	22.8	73.5	153873	9 AC026458	AC026458 Homo sapi
C 9	22	71.0	597	9 AY056037	AY056037 Homo sapi
C 10	22	68.4	67767	2 AC068480	AC068480 Homo sapi
C 11	21.2	68.4	185829	2 AC127182	AC127182 Rattus no
C 12	21.2	67.7	3255	8 AF348621	AF348621 Emericell
C 13	21	67.7	137615	2 AP005284	AP005284 Oryza sat
C 14	20.6	66.5	15083	2 AC112349	AC112349 Rattus no
C 15	20.6	66.5	169264	2 AC112889	AC112889 Rattus no
C 16	20.6	66.5	191557	9 AC008506	AC008506 Homo sapi
C 17	20.4	65.8	151283	2 AC102763	AC102763 Mus muscu
C 18	20.4	65.8	152470	9 AL135923	AL135923 Human DNA
C 19	20.4	65.8	157692	10 AL672177	AL672177 Mouse DNA
C 20	20.4	65.8	174704	2 AC122577	AC122577 Rattus no
C 21	20.4	65.8	186989	2 AC101913	AC101913 Mus muscu
C 22	20.4	65.8	219730	2 AC116661	AC116661 Mus muscu
C 23	20.4	65.2	158450	2 AC068082	AC068082 Homo sapi
C 24	20.2	65.2	162419	2 AC093737	AC093737 Homo sapi
C 25	20.2	65.2	192396	9 AC012668	AC012668 Homo sapi
C 26	20.2	65.2	347162	10 AC084070	AC084070 Mus muscu
C 27	20.2	64.5	25100	3 U49946	U49946 Caenorhabdi
C 28	20	64.5	68206	2 AC124323	AC124323 Mus muscu
C 29	20	64.5	70527	2 AC121560	AC121560 Mus muscu
C 30	20	64.5	110000	2 AC013622_3	Continuation (4 of
C 31	20	64.5	138783	9 HS564M11	AL035409 Human DNA
C 32	20	64.5	142422	9 AC113935	AC113935 Homo sapi
C 33	20	64.5	146072	2 AC091233	AC091233 Oryza sat
C 34	20	64.5	158145	2 AC128627	AC128627 Rattus no
C 35	20	64.5	161179	2 AC066601	AC066601 Homo sapi
C 36	20	64.5	171296	9 CNS06C86	AL391262 Human chr
C 37	20	64.5	173663	2 AC096960	AC096960 Rattus no
C 38	20	64.5	180366	9 AC036196	AC036196 Homo sapi
C 39	20	64.5	189540	10 AL772264	AL772264 Mouse DNA
C 40	20	64.5	193121	2 AC119482	AC119482 Rattus no
C 41	20	64.5	193865	2 AL845164	AL845164 Mus muscu
C 42	20	64.5	194237	9 AC068870	AC068870 Homo sapi
C 43	20	64.5	194487	2 AC127043	AC127043 Rattus no
C 44	20	64.5	194492	9 AC097625	AC097625 Homo sapi
C 45	20	64.5	194492	9 AC097625	AC097625 Homo sapi

## ALIGNMENTS

RESULT 1  
AB076888 774 bp mRNA linear PRI 25-JUN-2002  
LOCUS Homo sapiens mRNA for Di-Ras1, complete cds.  
DEFINITION AB076888  
ACCESSION AB076888  
VERSION AB076888.1 GI:21624247  
KEYWORDS  
SOURCE  
ORGANISM  
Homo sapiens  
brain cDNA.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE  
1 Kontani,K., Tada,M., Ogawa,T., Okai,T. and Katada,T.  
Di-Ras: A Distinct Subgroup of Ras-Family GTPases with Unique

Pred. No. is the number of results predicted by chance to have a

JOURNAL Biochemical Properties  
REFERENCE 2 (bases 1 to 774)  
AUTHORS Kontani,K., Ogawa,T., Okai,T., Tada,M. and Katada,T.  
TITLE Direct Submission  
JOURNAL Submitted (24-DEC-2001) Toshiaki Katada, University of Tokyo,  
Department of Physiological Chemistry, Graduate School of  
Pharmaceutical Sciences; 7-3-1 Honjo, Bunkyo-ku, Tokyo 113-0033,  
Japan (E-mail:katada@mol.f.u-tokyo.ac.jp, Tel:81-3-5841-4750,  
Fax:81-3-5841-4751)

FEATURES  
source Location/Qualifiers

1..774  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="19"  
/map="19p13.3"  
/tissue\_type="brain"  
/clone\_lib="lambda ZAPII human brain cDNA"  
/dev\_stage="adult"  
1..774  
/gene="DI-ras1"  
100..696  
/gene="DI-ras1"  
/note="member of the Ras family  
small GTP-binding protein"  
/codon\_start=1  
/product="DI-Ras1"  
/protein\_id="BAC01115.1"  
/db\_xref="GI:21624248"  
/translation="MPEOSNDYRVAVFGAGVGKSSLYLREVGKGFPRPTVPTIEDTY  
RQVISCCKSVCTLTDTGTSHPAPMRSLISGHAFILVFSYTSKSLDELPIRYK  
LIVQKSGVEDIPMLVGNKCDERQREYDTREAOVAQDEMCAFMETSAKMNVKEL  
FQELITETRNMSLINDGKRSQKQRTDRVKGCTLM"

BASE COUNT 154 a 268 c 239 g 113 t

Query Match 74.2%; Score 23; DB 9; Length 774;  
Best Local Similarity 100.0%; Pred. No. 2.1;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 9 CTCACATGAGGGTGCATTGGCCC 31  
|||||  
Db 697 CTCACATGAGGGTGCATTGGCCC 675

RESULT 2  
AX430295/c AX430295 1249 bp DNA linear PAT 28-JUN-2002  
DEFINITION Sequence 77 from Patent WO0240715.  
LOCUS AX430295  
VERSION AX430295.1 GI:21655659  
KEYWORDS human.  
SOURCE

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1  
AUTHORS Chalup,M.S., Altus,C.M., Lincoln,S.E., Dufour,G.E. and Jackson,S.  
TITLE Molecules for disease detection and treatment  
JOURNAL Patent: WO 0240715-A 77 23-MAY-2002;  
INCYTE GENOMICS INC (US)  
FEATURES  
source Location/Qualifiers

1..1249  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/note="Incyte ID No: LG:389870.1:2000SEP08"  
BASE COUNT 230 a 436 c 355 g 228 t

Query Match 74.2%; Score 23; DB 6; Length 1249;  
Best Local Similarity 100.0%; Pred. No. 2.1;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 CTCACATGAGGGTGCATTGGCCC 31  
|||||  
Db 647 CTCACATGAGGGTGCATTGGCCC 625

RESULT 3  
AK096600/c AK096600 2827 bp mRNA linear PRI 15-JUL-2002  
LOCUS Homo sapiens cDNA FLJ39281 fls. clone OCBBF2011067, highly similar  
DEFINITION to Homo sapiens cingulin mRNA.  
ACCESSION AK096600.1 GI:21756131  
VERSION oligo capping; fls (full insert sequence).  
KEYWORDS Homo sapiens fetal brain cDNA to mRNA, clone\_1lb:OCBBF2  
SOURCE clone:OCBBF2011067.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1  
AUTHORS Nishii,T., Ota,T., Nakagawa,S., Senoh,A., Mizuguchi,H., Inagaki,H.,  
Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,  
Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,  
Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,  
Kiyuchi,H., Kanda,K., Nagatsuna,M., Murakawa,K., Kanehori,K.,  
Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y.,  
Sugano,S., Nagahara,K., Masuho,Y., Nagai,K. and Isogai,T.  
NEO human cDNA sequencing project  
Unpublished

TITLE 2 (bases 1 to 2827)  
JOURNAL Isogai,T. and Yamamoto,J.  
REFERENCE Direct Submission  
AUTHORS Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team); 2-6-7  
JOURNAL Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan  
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)  
NEO human cDNA sequencing project supported by Ministry of  
Economy, Trade and Industry of Japan; cDNA full insert sequencing:  
Research Association for Biotechnology (RAB); cDNA library  
construction: Helix Research Institute (HRI) (supported by Japan  
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,  
HRI, and Biotechnology Center, National Institute of Technology and  
Evaluation; clone selection for full insert sequencing: HRI and  
RAB; annotation: HRI and RAB.

FEATURES  
source Location/Qualifiers

1..2827  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="OCBBF2011067"  
/tissue\_type="brain"  
/clone\_lib="OCBBF2"  
/dev\_stage="fetal"  
/note="cloning vector: PME18SFL3"  
BASE COUNT 688 a 750 c 983 g 406 t

Query Match 74.2%; Score 23; DB 9; Length 2827;  
Best Local Similarity 100.0%; Pred. No. 2.3;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 9 CTCACATGAGGGTGCATTGGCCC 31  
|||||  
Db 727 CTCACATGAGGGTGCATTGGCCC 705

RESULT 4  
AX430418/c AX430418 3061 bp DNA linear PAT 28-JUN-2002  
DEFINITION Sequence 200 from Patent WO0240715.  
LOCUS AX430418  
ACCESSION AX430418  
VERSION AX430418.1 GI:21655782  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 Mammalia; Eutheria; Primates; Catarrhini; Homiinae; Homo.  
AUTHORS Chaijap,M.S., Altus,C.M., Lincoln,S.E., Dufour,G.E. and Jackson,S.  
TITLE Molecules for disease detection and treatment  
JOURNAL Patent: WO 0240715-A 200 23-MAY-2002;  
INCYTE GENOMICS INC (US)  
FEATURES location/Qualifiers  
source 1. 3061  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/note="Incyte ID No: LI:332161.1:2000SEP08"  
BASE COUNT 606 a 840 c 927 g 688 t  
ORIGIN  
Query Match 74.2%; Score 23; DB 6; Length 3061;  
Best Local Similarity 100.0%; Pred. No. 2.3;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 9 CTCACATGAGGTCATTTGCC 31  
b 649 CTCACATGAGGTCATTTGCC 627  
RESULT 5 BC030660 3391 bp mRNA linear PRI 21-MAY-2002  
LOCUS Homo sapiens, similar to Rlg protein, clone MGC:33391  
DEFINITION IMAGE:4814337, mRNA, complete cds.  
ACCESSION BC030660  
VERSION BC030660.1 GI:21040534  
KEYWORDS MGC.  
SOURCE Homo sapiens.  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiinae; Homo.  
REFERENCE 1 (bases 1 to 3391)  
AUTHORS Strausberg,R.  
TITLE Direct Submission  
JOURNAL Submitted (20-MAY-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
COMMENT Contact: MGC help desk  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki  
Toshiyuki and Piero Carninci (RIKEN)  
DNA Sequencing by: Institute for Systems Biology  
<http://www.systemsbio.org>  
contact: amadan@systemsbio.org  
Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha  
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whitting  
FEATURES  
source  
1. 3391  
location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/db\_xref="taxon:9606"  
/clone="MGC:33391 IMAGE:4814337"  
/tissue\_type="Brain, hippocampus"  
/clone\_lib="NIH\_MGC\_95"  
/lab\_host="DH10B"  
/note="Vector: pBluescript"  
/codon\_start=1  
CDS

/product="similar to Rlg protein"  
/protein\_id="AAH30660.1"  
/db\_xref="GI:21040535"  
/translation="MPROSDRYRVYVAGGVGSSVLRFVKGTFRDPYIPIYEDRY  
ROYISCKSVCTLOITDTTSSHOPFAPMORISIKGHAFILVFSTYSKOSIPELAPIYK  
LIVQIKGSVDIDPVMALVNCDETRVEDYREAQVAQEWKCAFMETSARKNVKEL  
FOELTLETFRRNMLNIDGRSGRKQKTDVKGKCTLM"  
BASE COUNT 650 a 996 c 1016 g 729 t  
ORIGIN  
Query Match 74.2%; Score 23; DB 9; Length 3391;  
Best Local Similarity 100.0%; Pred. No. 2.3;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 9 CTCACATGAGGTCATTTGCC 31  
Db 735 CTCACATGAGGTCATTTGCC 713  
RESULT 6 AC006538 177540 bp DNA linear PRI 07-FEB-1999  
LOCUS Homo sapiens chromosome 19, BAC 41195 (CIT-B-31c16), complete  
DEFINITION sequence.  
ACCESSION AC006538  
VERSION AC006538.1 GI:4235145  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
REFERENCE 1 (bases 1 to 177540)  
AUTHORS Lamerdin,J.E., McCreedy,P.M., Skowronski,E., Viswanathan,V.,  
Burkhardt-Schultz,K., Gordon,L., Dias,J., Ramirez,M., Stillwagen,S.,  
Phan,H., Velasco,N., Do,L., Regala,W., Terry,A., Ganes,J.,  
Danganan,L., Erler,A., Christensen,M., Georgescu,A., Avila,J.,  
Liu,S., Altix,C., Andreise,T., Frankheim,M., Amico-Keller,G.,  
Coefield,J., Duarte,S., Lucas,S., Bruce,R., Thomas,P., Quan,G.,  
Krommiller,B., Arellano,A., Sanders,C., Ow,D., Nolan,M., Trong,S.,  
Kobayashi,A., Olsen,A.S. and Carrano,A.V.  
TITLE Sequence analysis of a 3.5 Mb contig in human 19p13.3 containing a  
serine protease gene cluster  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 177540)  
AUTHORS Lamerdin,J.E.  
TITLE Direct Submission  
JOURNAL Submitted (07-FEB-1999) Joint Genome Institute, Lawrence Livermore  
National Laboratory, 7000 East Ave., Livermore, CA 94551, USA  
COMMENT Map and sequence oriented from p telomere to centromere. BC41195  
(CIT-B-31c16) is currently separated from cosmid R32203 (AC006275)  
to the left by a sequence gap of approximately 6 kb, and overlaps  
BAC 102889 (CIT-B-191b6: AC006130) to the right from bases 175,545  
to 177,540. Additional map and sequence information are available  
at: <http://www.bio.illn.gov/db/rgp/genome/genome.html>.  
FEATURES  
source  
1. 177540  
location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="19"  
/map="19p13.3 between CDC34 and D19S342"  
/clone="CIT-B-31c16"  
/clone\_lib="Cal tech CIT-B BAC library"  
/note="LNM clone name: BC41195"  
complement(1..144)  
/rpt\_family="LIMC/D"  
283..381  
/rpt\_family="AluJo"  
complement(399..437)  
/rpt\_family="(GAA)n"  
complement(438..738)  
/rpt\_family="AluSx"  
complement(868..931)  
/rpt\_family="LIMB7"  
repeat\_region  
repeat\_region  
repeat\_region  
repeat\_region  
repeat\_region

```
misc_feature      1175..1365  
/note="predicted exon, program: grail2exons_human_1.3,  
frame: 2, quality: good, score: 57.000"  
repeat_region    /rpt_family="AluSg"  
complement(2133..2231)  
repeat_region    /rpt_family="MER81"  
2739..2935  
repeat_region    /rpt_family="AluJo"  
2946..3247  
repeat_region    /rpt_family="AluSx"  
3256..3384  
repeat_region    /rpt_family="L1MC3"  
3389..3604  
repeat_region    /rpt_family="AluJo"  
3631..3716  
repeat_region    /rpt_family="L1MD3"  
complement(4087..4242)  
repeat_region    /rpt_family="AluJo"  
complement(4294..4436)  
repeat_region    /rpt_family="AluJo"  
complement(4598..4887)  
repeat_region    /rpt_family="AluSg"  
complement(4906..5373)  
repeat_region    /rpt_family="L1MB7"  
complement(5382..5669)  
repeat_region    /rpt_family="AluSx"  
complement(6425..6609)  
repeat_region    /rpt_family="(GGA)n"  
complement(6610..6765)  
repeat_region    /rpt_family="(GAA)n"  
complement(6766..7065)  
repeat_region    /rpt_family="AluSx"  
7093..7392  
repeat_region    /rpt_family="AluSx"  
7423..7511  
repeat_region    /rpt_family="L1MA7"  
7539..7839  
repeat_region    /rpt_family="AluSx"  
7852..7938  
repeat_region    /rpt_family="L1MA7"  
8001..8312  
repeat_region    /rpt_family="AluJo"  
complement(8771..8897)  
repeat_region    /rpt_family="MER20"  
complement(9532..9722)  
repeat_region    /rpt_family="MER20"  
complement(9970..10143)  
repeat_region    /rpt_family="MER20"  
10144..10405  
repeat_region    /rpt_family="AluJo"  
10421..10630  
repeat_region    /rpt_family="L1M3"  
10631..10756  
repeat_region    /rpt_family="AluSg"  
10757..11058  
repeat_region    /rpt_family="AluY"  
11068..11243  
repeat_region    /rpt_family="AluSg"  
11247..11520  
repeat_region    /rpt_family="L1M3"  
11522..11816  
repeat_region    /rpt_family="AluSg"  
complement(11827..12033)  
repeat_region    /rpt_family="MER58A"  
12255..12555  
repeat_region    /rpt_family="AluJo"  
complement(12562..12617)  
repeat_region    /rpt_family="(TA)n"  
12769..12894  
repeat_region    /rpt_family="FLAM_A"  
complement(12901..12951)  
repeat_region    /rpt_family="L1MC3"  
  
repeat_region    12960..13013  
repeat_region    /rpt_family="LFR13"  
13013..13408  
repeat_region    /rpt_family="LFR13"  
complement(13409..13457)  
repeat_region    /rpt_family="L1MD3"  
13651..13908  
repeat_region    /rpt_family="L1MB8"  
13906..14116  
repeat_region    /rpt_family="L1MC2"  
complement(14173..14468)  
repeat_region    /rpt_family="AluY"  
14497..14811  
repeat_region    /rpt_family="AluSx"  
14812..15011  
repeat_region    /rpt_family="(GGA)n"  
15689..16007  
repeat_region    /rpt_family="AluJo"  
16008..16128  
repeat_region    /rpt_family="MLT1F"  
16130..16430  
repeat_region    /rpt_family="AluSp"  
16432..16563  
repeat_region    /rpt_family="(GAA)n"  
16575..16891  
repeat_region    /rpt_family="AluSx"  
16892..16930  
repeat_region    /rpt_family="(GAA)n"  
16989..17073  
repeat_region    /rpt_family="MLT1F"  
17178..17479  
repeat_region    /rpt_family="AluSx"  
17481..18206  
repeat_region    /rpt_family="MER21B"  
18240..18329  
repeat_region    /rpt_family="MLT1F"  
18380..18748  
repeat_region    /rpt_family="MSTD"  
complement(19030..19229)  
repeat_region    /rpt_family="MER20"  
complement(19579..19692)  
misc_feature    /note="predicted exon, program: grail2exons_human_1.3,  
frame: 1, quality: good, score: 53.000"  
repeat_region    /rpt_family="MER20"  
complement(20388..20585)  
repeat_region    /rpt_family="MER20"  
20846..20994  
repeat_region    /rpt_family="MIR"  
complement(21068..21332)  
repeat_region    /rpt_family="AluSx"  
complement(21372..21671)  
repeat_region    /rpt_family="AluSg"  
complement(21681..21811)  
repeat_region    /rpt_family="FLAM_C"  
21837..22056  
repeat_region    /rpt_family="MER58A"  
22240..22539  
repeat_region    /rpt_family="AluSx"  
22697..22831  
repeat_region    /rpt_family="AluJo"  
22833..23132  
repeat_region    /rpt_family="AluSx"  
23136..23412  
repeat_region    /rpt_family="L1MC3"  
complement(24002..24172)  
repeat_region    /rpt_family="MER20"  
complement(24191..24313)  
repeat_region    /rpt_family="MIR"  
24410..24707  
repeat_region    /rpt_family="AluSg"  
complement(24739..24926)  
repeat_region    /rpt_family="MER58A"
```

```

repeat_region 24928 .25229
                /rpt_family="Alusx"

Query Match      74.28; Score 23; DB 9; Length 177540;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 9 CTCACATGAGGTCGTCATTGCC 31
Db 72409 CTCACATGAGGTCGTCATTGCC 72431

RESULT 7
AC113495 194456 bp DNA linear HTG 06-JUN-2002
LOCUS Mus musculus clone RP23-361011, WORKING DRAFT SEQUENCE, 14 ordered
DEFINITION pieces.
AC113495 GI:21327431
VERSION HTG: HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 194456)
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
JOURNAL Unpublished
TITLE 2 (bases 1 to 194456)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhalter, B.,
Brown, A., Camarata, N., Campopiano, A., Chang, J., Chazaro, B.,
Choepel, Y., Collangelo, M., Collins, S., Collamore, A., Cook, A.,
Cooke, P., Dearliano, K., Dewar, K., Diaz, J. S., Dodge, S., Fato, S.,
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Labrecque, K., Lamazares, R.,
Landers, T., Lehoczeky, J., Levine, R., Liu, G., Maclean, C.,
Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M.,
McEwan, P., McKernan, K., Meldrim, J., Menes, L., Mihova, T.,
Mlenka, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,
Rett, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,
Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S.,
Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (01-MAR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 194456)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,
Boukhalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
Chazaro, B., Choepel, Y., Collangelo, M., Collins, S., Collamore, A.,
Cook, A., Cooke, P., Dearliano, K., Dewar, K., Diaz, J. S., Dodge, S.,
Dodge, S., Ferreira, P., Fitzgerald, M., Fitzhugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I.,
Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Labrecque, K.,
Lamazares, R., Landers, T., Lehoczeky, J., Levine, R., Lindblad-Toh, K.,
Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N.,
Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J.,
Menes, L., Mihova, T., Mlenka, V., Murphy, T., Naylor, J., Nguyen, C.,
Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
Pollara, V., Raymond, C., Rett, R., Rieback, M., Riley, R., Rise, C.,
Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S.,
Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N.,

```

```

TITLE
JOURNAL
COMMENT
Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S.,
Theodore, J., Topham, K., Travers, M., Travis, N., Triggillo, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (06-JUN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 6, 2002 this sequence version replaced gi:21313668.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-----
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-----
Project Information
Center project name: L23814
Center clone name: 361_O_11
-----
Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 189249 bases at least Q40
Consensus quality: 191283 bases at least Q30
Consensus quality: 192097 bases at least Q20
Insert size: 194000; agarose-fp
Insert size: 193156; sum-of-consigs
Quality coverage: 7.8 in Q20 bases; sum-of-consigs
Quality coverage: 7.9 in Q20 bases; sum-of-consigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
1 320: contig of 320 bp in length
321 420: gap of 100 bp
421 1287: contig of 867 bp in length
1288 1387: gap of 100 bp
1388 3055: contig of 1668 bp in length
3056 3155: gap of 100 bp
3156 4819: contig of 1664 bp in length
4820 4919: gap of 100 bp
4920 7593: contig of 2674 bp in length
7594 7693: gap of 100 bp
7694 14544: contig of 6851 bp in length
14545 14644: gap of 100 bp
14645 21443: contig of 6799 bp in length
21444 21543: gap of 100 bp
21544 34699: contig of 13156 bp in length
34700 34799: gap of 100 bp
34800 51238: contig of 16439 bp in length
51239 51338: gap of 100 bp
51339 70476: contig of 19138 bp in length
70477 70576: gap of 100 bp
70577 123829: contig of 53253 bp in length
123830 123929: gap of 100 bp
123930 147950: contig of 24021 bp in length
147951 148050: gap of 100 bp
148051 179165: contig of 31115 bp in length
179166 179265: gap of 100 bp
179266 194456: contig of 15191 bp in length.
-----
FEATURES
SOURCE
location/Qualifiers
1..194456
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-361011"
/clone_lib="RPCI-23 Female Mouse BAC"

```

```
misc_feature 1..320
/note="assembly_fragment
clone_end:SP6
vector_side:left"
misc_feature 421..1287
/note="assembly_fragment"
1388..3055
/note="assembly_fragment"
3156..4819
/note="assembly_fragment"
4920..7593
/note="assembly_fragment"
7694..14344
/note="assembly_fragment"
14645..21443
/note="assembly_fragment"
21544..34699
/note="assembly_fragment"
34800..51238
/note="assembly_fragment"
51339..70476
/note="assembly_fragment"
70577..123829
/note="assembly_fragment"
123930..147950
/note="assembly_fragment"
148051..179165
/note="assembly_fragment"
179266..194456
/note="assembly_fragment
clone_end:T7
vector_side:right"
BASE COUNT 52225 a 42789 c 44383 g 53744 t 1315 others
ORIGIN
Query Match 74.2% Score 23; DB 2; Length 194456;
Best Local Similarity 83.9% Pred. No. 3.1;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 GCGGAATTCACATGAGGTCATTGGCCC 31
11111111111111111111111111111111
Db 53017 GGTCAATTCCTGAGTACAGGTGCTACGACCC 53047
RESULT 8
AC026458 153873 bp DNA linear PRI 30-MAR-2002
LOCUS Homo sapiens chromosome 16 clone RP11-10K17, complete sequence.
DEFINITION AC026458
AC026458.7 GI:19848340
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 153873)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 153873)
TITLE DOE Joint Genome Institute.
JOURNAL Direct Submission
AUTHORS 3 (bases 1 to 153873)
TITLE Production Sequencing Facility, DOE Joint
JOURNAL Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 4 (bases 1 to 153873)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (30-MAR-2002) Production Sequencing Facility, DOE Joint
```

```
COMMENT
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Mar 30, 2002 this sequence version replaced gi:18767414.
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
Genbank flat file format but are available as part
of this entry's ASN.1 file.
-----
Sequence quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
Genbank flat file format but are available as part
of this entry's ASN.1 file.
-----
FEATURES
Source Location/Qualifiers
1..153873
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-10K17"
BASE COUNT 37256 a 37090 c 39282 g 40245 t
ORIGIN
Query Match 73.5% Score 22.8; DB 9; Length 153873;
Best Local Similarity 92.3% Pred. No. 3.9;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 6 ATTCTCACATGAGGTCATTGGCCC 31
11111111111111111111111111111111
Db 132268 ATTCTCAATGAGGTCATTGGCCC 132293
RESULT 9
AY056037 597 bp mRNA linear PRI 24-JUL-2002
LOCUS Homo sapiens Rlg protein mRNA, complete cds.
DEFINITION AY056037
ACCESSION AY056037.1 GI:16555333
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 597)
AUTHORS Ellis,C.A., Vos,M.D., Howell,H., Vallecorsa,T., Fults,D.W. and
Clark,G.J.
TITLE Rlg is a novel Ras-related protein and potential neural tumor
suppressor
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (15), 9876-9881 (2002)
AUTHORS 2 (bases 1 to 597)
TITLE Castro,A.F. and Quilliam,L.A.
AUTHORS Identification of Rlg, a novel Ras family member sharing homology
with NOY2 and Rap
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 597)
AUTHORS Castro,A.F. and Quilliam,L.A.
TITLE Direct Submission
JOURNAL Submitted (13-SEP-2001) Biochemistry and Molecular Biology, Indiana
University School of Medicine, 635 Barnhill Drive, MS-4053,
Indianapolis, IN 46202, USA
FEATURES
Source Location/Qualifiers
1..597
/organism="Homo sapiens"
/db_xref="taxon:9606"
```

```

CDS
1..597
/feature="Ras family GTPase"
/codon_start=1
/product="R1g protein"
/protein_id="AAL23715.1"
/db_xref="GI:1655334"
/translation="MEQSDNYRVYVFGAGVGKSSLVLFKVGTFRDYIPIEDTY
ROYISCDKSVCTQITDTTGSHPAPMORLISKGAFILVSVTSKQSLSEGLPIYK
LIVQISGVEDIVMLVGNKCDREVDREAOVAQOEKCAFMTSAAKMNYNVKEL
FOELLTETFRNNSLINDGKRSKOKRTDVKKCTILM"

BASE COUNT      134 a      178 c      195 g      90 t
ORIGIN
Query Match      71.0%; Score 22; DB 9; Length 597;
Best Local Similarity 100.0%; Pred. No. 6.6;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y      10 TCACATGAGGTCATTGCC 31
597 TCACATGAGGTCATTGCC 576

RESULT 10
AY059641      597 bp mRNA linear PRI 29-OCT-2001
LOCUS
DEFINITION
Homo sapiens small GTP-binding tumor suppressor 1 mRNA, complete
cds.
ACCESSION
AY059641
VERSION
AY059641.1 GI:16508175
KEYWORDS
Homo sapiens.
ORGANISM
Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthelia; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 597)
Gong, L. and Wu, K.
Molecular cloning of GBRSL, a novel gene encoding a small
GTP-binding tumor suppressor
Unpublished
2 (bases 1 to 597)
Gong, L.
Direct Submission
Submitted (12-OCT-2001) Dept. of Cardiology, U.T. MD. Anderson
Cancer Center, 1515 Holcombe Blvd-449, Houston, TX 77030, USA
Location/Qualifiers
1..597
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
1..597
/note="GBRSL"
/codon_start=1
/product="small GTP-binding tumor suppressor 1"
/protein_id="AAL17968.1"
/db_xref="GI:16508176"
/translation="MEQSDNYRVYVFGAGVGKSSLVLFKVGTFRDYIPIEDTY
ROYISCDKSVCTQITDTTGSHPAPMORLISKGAFILVSVTSKQSLSEGLPIYK
LIVQISGVEDIVMLVGNKCDREVDREAOVAQOEKCAFMTSAAKMNYNVKEL
FOELLTETFRNNSLINDGKRSKOKRTDVKKCTILM"

BASE COUNT      134 a      178 c      195 g      90 t
ORIGIN
Query Match      71.0%; Score 22; DB 9; Length 597;
Best Local Similarity 100.0%; Pred. No. 6.6;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y      10 TCACATGAGGTCATTGCC 31
597 TCACATGAGGTCATTGCC 576

```

```

RESULT 11
AC068480
LOCUS
DEFINITION
Homo sapiens chromosome 3 clone RP11-365F21 map 3, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION
AC068480
VERSION
AC068480.1 GI:767964
KEYWORDS
HTG; HTGS-PHASED.
SOURCE
Homo sapiens.
ORGANISM
Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthelia; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 67767)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 3, clone RP11-365F21
Unpublished
2 (bases 1 to 67767)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bada, F.,
Boguslavsky, L., Bouckhalter, B., Brown, A., Burkett, G.,
Campopiano, A., Castle, A., Choquet, Y., Colangelo, K., Collins, S.,
Collins, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,
Dodgson, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., Lacroque, K., Lamazares, R., Landers, J., Lehotzky, J.,
Levine, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
McLarty, J., Menus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rotman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tessier, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (02-MAY-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smith, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: U990
Center clone name: 365_F_21
-----
* NOTE: This record contains 80 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1 757: contig of 757 bp in length
* 758 857: gap of 100 bp
* 858 1628: contig of 771 bp in length
* 1629 1728: gap of 100 bp
* 1729 2483: contig of 755 bp in length
* 2484 2583: gap of 100 bp
* 2584 3339: contig of 756 bp in length

```

\* 3340 3439: gap of 100 bp  
\* 3440 4197: contig of 758 bp in length  
\* 4198 4297: gap of 100 bp  
\* 4298 5035: contig of 738 bp in length  
\* 5036 5135: gap of 100 bp  
\* 5136 5864: contig of 729 bp in length  
\* 5865 5964: gap of 100 bp  
\* 5965 6700: contig of 736 bp in length  
\* 6701 6800: gap of 100 bp  
\* 6801 7544: contig of 744 bp in length  
\* 7545 7644: gap of 100 bp  
\* 7645 8404: contig of 760 bp in length  
\* 8405 8504: gap of 100 bp  
\* 8505 9238: contig of 734 bp in length  
\* 9239 9338: gap of 100 bp  
\* 9339 10092: contig of 754 bp in length  
\* 10093 10192: gap of 100 bp  
\* 10193 10956: contig of 764 bp in length  
\* 10957 11056: gap of 100 bp  
\* 11057 11609: contig of 753 bp in length  
\* 11810 11909: gap of 100 bp  
\* 11910 12658: contig of 749 bp in length  
\* 12659 12758: gap of 100 bp  
\* 12759 13505: contig of 747 bp in length  
\* 13506 13605: gap of 100 bp  
\* 13606 14343: contig of 738 bp in length  
\* 14344 14443: gap of 100 bp  
\* 14444 15177: contig of 734 bp in length  
\* 15178 15277: gap of 100 bp  
\* 15278 16020: contig of 743 bp in length  
\* 16021 16120: gap of 100 bp  
\* 16121 16874: contig of 754 bp in length  
\* 16875 16974: gap of 100 bp  
\* 16975 17719: contig of 745 bp in length  
\* 17720 17819: gap of 100 bp  
\* 17820 18558: contig of 739 bp in length  
\* 18559 18658: gap of 100 bp  
\* 18659 19414: contig of 756 bp in length  
\* 19415 19514: gap of 100 bp  
\* 19515 20260: contig of 746 bp in length  
\* 20261 20360: gap of 100 bp  
\* 20361 21109: contig of 749 bp in length  
\* 21110 21209: gap of 100 bp  
\* 21210 21959: contig of 750 bp in length  
\* 21960 22059: gap of 100 bp  
\* 22060 22808: contig of 749 bp in length  
\* 22809 22908: gap of 100 bp  
\* 22909 23660: contig of 752 bp in length  
\* 23661 23760: gap of 100 bp  
\* 23761 24517: contig of 757 bp in length  
\* 24518 24617: gap of 100 bp  
\* 24618 25378: contig of 761 bp in length  
\* 25379 25478: gap of 100 bp  
\* 25479 26227: contig of 749 bp in length  
\* 26228 26327: gap of 100 bp  
\* 26328 27083: contig of 756 bp in length  
\* 27084 27183: gap of 100 bp  
\* 27184 27929: contig of 746 bp in length  
\* 27930 28029: gap of 100 bp  
\* 28030 28778: contig of 749 bp in length  
\* 28779 28878: gap of 100 bp  
\* 28879 29642: contig of 764 bp in length  
\* 29643 29742: gap of 100 bp  
\* 29743 30497: contig of 755 bp in length  
\* 30498 30597: gap of 100 bp  
\* 30598 31347: contig of 750 bp in length  
\* 31348 31447: gap of 100 bp  
\* 31448 32194: contig of 747 bp in length  
\* 32195 32294: gap of 100 bp  
\* 32295 33047: contig of 753 bp in length  
\* 33048 33147: gap of 100 bp  
\* 33148 33879: contig of 732 bp in length  
\* 33880 33979: gap of 100 bp

\* 33980 34731: contig of 752 bp in length  
\* 34732 34831: gap of 100 bp  
\* 34832 35583: contig of 752 bp in length  
\* 35584 35683: gap of 100 bp  
\* 35684 36423: contig of 740 bp in length  
\* 36424 36523: gap of 100 bp  
\* 36524 37258: contig of 735 bp in length  
\* 37259 37358: gap of 100 bp  
\* 37359 38110: contig of 752 bp in length  
\* 38111 38210: gap of 100 bp  
\* 38211 38957: contig of 747 bp in length  
\* 38958 39057: gap of 100 bp  
\* 39058 39608: contig of 751 bp in length  
\* 39609 39908: gap of 100 bp  
\* 39909 40663: contig of 755 bp in length  
\* 40664 40763: gap of 100 bp  
\* 40764 41521: contig of 758 bp in length  
\* 41522 41621: gap of 100 bp  
\* 41622 42359: contig of 738 bp in length  
\* 42360 42459: gap of 100 bp  
\* 42460 43193: contig of 734 bp in length  
\* 43194 43293: gap of 100 bp  
\* 43294 44028: contig of 735 bp in length  
\* 44029 44128: gap of 100 bp  
\* 44129 44891: contig of 763 bp in length  
\* 44892 44991: gap of 100 bp  
\* 44992 45752: contig of 761 bp in length  
\* 45753 45852: gap of 100 bp  
\* 45853 46593: contig of 741 bp in length  
\* 46594 46693: gap of 100 bp  
\* 46694 47441: contig of 748 bp in length  
\* 47442 47541: gap of 100 bp  
\* 47542 48288: contig of 747 bp in length  
\* 48289 48388: gap of 100 bp  
\* 48389 49130: contig of 742 bp in length  
\* 49131 49230: gap of 100 bp  
\* 49231 49980: contig of 750 bp in length  
\* 49981 50080: gap of 100 bp  
\* 50081 50827: contig of 747 bp in length  
\* 50828 50927: gap of 100 bp  
\* 50928 51669: contig of 742 bp in length  
\* 51670 51769: gap of 100 bp  
\* 51770 52512: contig of 743 bp in length  
\* 52513 52612: gap of 100 bp  
\* 52613 53352: contig of 740 bp in length  
\* 53353 53452: gap of 100 bp  
\* 53453 54194: contig of 742 bp in length  
\* 54195 54294: gap of 100 bp  
\* 54295 55049: contig of 755 bp in length  
\* 55050 55149: gap of 100 bp  
\* 55150 55896: contig of 747 bp in length  
\* 55897 55996: gap of 100 bp  
\* 55997 56750: contig of 754 bp in length  
\* 56751 56850: gap of 100 bp  
\* 56851 57604: contig of 754 bp in length  
\* 57605 57704: gap of 100 bp  
\* 57705 58411: contig of 707 bp in length  
\* 58412 58511: gap of 100 bp

Query Match 68.4%; Score 21.2; DB 2; Length 67767;  
Best Local Similarity 88.5%; Pred. No. 24;  
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 GCGAATTCACATGAGGATGCATT 26  
Db 27478 GCGAATTCACCTGAGGAGCGTT 27503

RESULT 12  
AC127182/c 185829 bp DNA linear HTG 14-JUL-2002  
LOCUS Rattus norvegicus clone CH230-96H21, \*\*\* SEQUENCING IN PROGRESS  
DEFINITION \*\*\*. 63 unordered pieces.



ACCESSION	AC127182
VERSION	AC127182.1
KEYWORDS	GI:21747570
SOURCE	HTG; HTGS_PHASE1.
ORGANISM	Norway rat. Rattus norvegicus
REFERENCE	1 (bases 1 to 185829)
AUTHORS	Muzny,D.M., Adams,C., Adio-oduola,B., Ali-osman,F.R., Allen,C., Alshrooks,S.L., Amaratunga,H.C., Are,J.R., Ayele,M., Banks,T., Barbakia,J., Benton,J., Bimaga,K., Blankenburg,K., Bonini,D., Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorielli,D.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Huliy,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratoch,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Louised,H., Lozad,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheswari,M., Mapa,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mel,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshitari,N., Sisson,I., Soederren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,O., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and gibbs,R.
TITLE	Direct Submission
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 185829)
AUTHORS	Worley,K.C.
TITLE	Direct Submission
JOURNAL	Submitted (14-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT	----- Genome Center Center: Baylor College of Medicine Center code: BCM Web site: <a href="http://www.hgsc.bcm.tmc.edu/">http://www.hgsc.bcm.tmc.edu/</a> Contact: hgsc-help@bcm.tmc.edu ----- Project Information Center project name: G00S Center clone name: CH230-96H21 ----- Summary Statistics Sequencing vector: Plasmid; Chemistry: Dye-terminator Big Dye; 100% of reads Assembly program: Phrap; version 0.990329 Consensus quality: 131591 bases at least Q40 Consensus quality: 137671 bases at least Q30 Consensus quality: 141923 bases at least Q20 -----

* NOTE: Estimated insert size may differ from sequence length
* (see <a href="http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html">http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html</a> ).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 63 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1083: contig of 1083 bp in length
1084
1183: gap of unknown length
1184
2406: contig of 1223 bp in length
2407
2506: gap of unknown length
4126: contig of 1620 bp in length
4127
4226: gap of unknown length
4227
5537: contig of 1311 bp in length
5538
5637: gap of unknown length
5638
7415: contig of 1778 bp in length
7416
7515: gap of unknown length
7516
8626: contig of 1111 bp in length
8627
8726: gap of unknown length
10003: contig of 1277 bp in length
10004
10103: gap of unknown length
10104
11880: contig of 1777 bp in length
11881
11980: gap of unknown length
11981
12985: contig of 1005 bp in length
12986
13085: gap of unknown length
13086
14457: contig of 1372 bp in length
14458
14557: gap of unknown length
14558
15698: contig of 1141 bp in length
15699
15798: gap of unknown length
15799
17182: contig of 1384 bp in length
17183
17282: gap of unknown length
17283
18534: contig of 1252 bp in length
18535
18634: gap of unknown length
18635
19953: contig of 1319 bp in length
19954
20053: gap of unknown length
20054
21317: contig of 1264 bp in length
21318
21417: gap of unknown length
21418
22841: contig of 1424 bp in length
22842
22941: gap of unknown length
22942
24544: contig of 1603 bp in length
24545
24644: gap of unknown length
24645
26066: contig of 1422 bp in length
26067
26166: gap of unknown length
26167
27351: contig of 1185 bp in length
27352
27451: gap of unknown length
27452
28608: contig of 1157 bp in length
28609
28708: gap of unknown length
28709
30940: contig of 2232 bp in length
30941
31040: gap of unknown length
31041
32702: contig of 1662 bp in length
32703
32802: gap of unknown length
32803
34379: contig of 1577 bp in length
34379
34479: gap of unknown length
34480
36044: contig of 1565 bp in length
36045
36144: gap of unknown length
36145
38735: contig of 2591 bp in length
38736
38835: gap of unknown length
38836
40281: contig of 1446 bp in length
40282
40381: gap of unknown length
40382
42272: contig of 1891 bp in length
42273
42372: gap of unknown length
42373
44841: contig of 2469 bp in length
44842
44941: gap of unknown length
44942
46361: contig of 1420 bp in length
46362
46461: gap of unknown length
46462
48242: contig of 1781 bp in length
48243
48342: gap of unknown length
48343
49344: contig of 1002 bp in length
49345
49444: gap of unknown length
49445
52526: contig of 3082 bp in length

```

* 52527 52526: gap of unknown length
* 52627 54749: contig of 2123 bp in length
* 54750 54849: gap of unknown length
* 54850 56762: contig of 1913 bp in length
* 56763 56862: gap of unknown length
* 56863 59578: contig of 2716 bp in length
* 59579 59678: gap of unknown length
* 59679 63989: contig of 4311 bp in length
* 63990 64089: gap of unknown length
* 64090 67207: contig of 3118 bp in length
* 67208 70821: gap of unknown length
* 70822 70921: gap of unknown length
* 70922 73810: contig of 2869 bp in length
* 73811 73910: gap of unknown length
* 73911 76613: contig of 2703 bp in length
* 76614 76713: gap of unknown length
* 76714 79312: contig of 2599 bp in length
* 79313 79412: gap of unknown length
* 79413 81825: contig of 2413 bp in length
* 81826 81925: gap of unknown length
* 81926 85896: contig of 3971 bp in length
* 85897 85996: gap of unknown length
* 85997 90144: contig of 4148 bp in length
* 90145 90244: gap of unknown length
* 90245 93598: contig of 3354 bp in length
* 93599 93698: gap of unknown length
* 93699 97232: contig of 3534 bp in length
* 97233 97332: gap of unknown length
* 97333 102574: contig of 5242 bp in length
* 102575 102674: gap of unknown length
* 102675 105003: contig of 2329 bp in length
* 105004 105103: gap of unknown length
* 105104 109204: contig of 4101 bp in length
* 109205 109304: gap of unknown length
* 109305 114272: contig of 4968 bp in length
* 114273 114372: gap of unknown length
* 114373 118712: contig of 4340 bp in length
* 118713 118812: gap of unknown length
* 118813 12381: contig of 4469 bp in length
* 12382 12381: gap of unknown length
* 12382 12751: contig of 4370 bp in length
* 12752 12751: gap of unknown length
* 12752 133414: contig of 5563 bp in length
* 133415 133514: gap of unknown length
* 133515 137510: contig of 3996 bp in length
* 137511 137610: gap of unknown length
* 137611 142739: contig of 5129 bp in length
* 142740 142839: gap of unknown length
* 142840 147223: contig of 4384 bp in length
* 147224 147323: gap of unknown length

```

Query Match 68.4%; Score 21.2; DB 2; Length 185829;  
 Best Local Similarity 88.5%; Pred. No. 26;  
 Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 3 GGAATTCGACATGAGGTCATTG 28  
 ||||| ||| ||||| ||||| |||||  
 Db 172402 GGAATTCGACATGAGGTCATTG 172377

RESULT 13  
 AF348621 3255 bp DNA linear PLN 03-APR-2001  
 LOCUS AF348621 unguis acetamidase (amds) gene, complete cds.  
 DEFINITION AF348621  
 ACCESSION AF348621  
 VERSION AF348621.1 GI:13517938  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Emericella unguis.  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 Eurotiales; Trichocomaceae; Emericella.  
 REFERENCE  
 1 (bases 1 to 3255)

AUTHORS Sharp, J.A., Davis, M.A. and Hynes, M.J.  
 TITLE Comparison of amidase genes in Aspergillus species  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 3255)  
 AUTHORS Sharp, J.A., Davis, M.A. and Hynes, M.J.  
 TITLE Direct Submission  
 JOURNAL Submitted (05-FEB-2001) Genetics, University of Melbourne, Royal Pde, Parkville, Vic 3010, Australia  
 FEATURES  
 source  
 1..3255  
 /organism="Emericella unguis"  
 /db\_xref="taxon:41737"  
 <946..>2781  
 /gene="amds"  
 join(<946..1203,1258..1398,1449..1869,1921..2078,  
 2125..>2781)  
 /gene="amds"  
 /product="acetamidase"  
 join(946..1203,1258..1398,1449..1869,1921..2078,  
 2125..>2781)  
 /gene="amds"  
 /codon\_start=1  
 /product="acetamidase"  
 /protein\_id="AAK29062.1"  
 /db\_xref="GI:13517939"

## CDS

BASE COUNT 769 a 852 c 820 g 814 t  
 ORIGIN  
 Query Match 67.7%; Score 21; DB 8; Length 3255;  
 Best Local Similarity 82.8%; Pred. No. 24;  
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 3 GGAATTCGACATGAGGTCATTGCCC 31  
 ||||| ||| ||||| ||||| |||||  
 Db 1272 GGAATTCGACATGAGGTCAGCTCGCC 1300

RESULT 14  
 AP005284 137615 bp DNA linear HTG 31-MAY-2002  
 LOCUS AP005284  
 DEFINITION Oryza sativa (japonica cultivar-group) chromosome 2 clone B112A12,  
 \*\*\* SEQUENCING IN PROGRESS \*\*\*. In ordered pieces.  
 ACCESSION AP005284  
 VERSION AP005284.1 GI:21280334  
 KEYWORDS HTG; HTGS; PHASE2.  
 SOURCE HTG; HTGS; PHASE2.  
 ORGANISM  
 Oryza sativa (japonica cultivar-group) (cultivar: Nipponbare) DNA,  
 clone: B112A12.  
 Oryza sativa (japonica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Eriatoidae; Oryzae; Oryza.  
 REFERENCE  
 1  
 Sasaki, T., Matsumoto, T. and Katayose, Y.  
 Oryza sativa nipponbare(GA3) genomic DNA, chromosome 2, BAC  
 clone: B112A12  
 Published Only in Database (2002)  
 2 (bases 1 to 137615)  
 Sasaki, T., Matsumoto, T. and Katayose, Y.  
 Direct Submission  
 Submitted (29-MAY-2002) Takuji Sasaki, National Institute of  
 Agrobiological Sciences, Rice Genome Research Program, Kannondai  
 2-1-2, Tsukuba, Ibaraki 305-8602, Japan  
 (E-mail: tsasaki@nias.affrc.go.jp, URL: http://irg.p.dna.affrc.go.jp/,

## COMMENT

Tel:81-298-38-7441, Fax:81-298-38-7468)  
 NOTE: It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have provided by the submitter. This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.  
 \* NOTE: This is a 'working draft' sequence.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.

## FEATURES

source  
 Location/Qualifiers  
 1. 137615  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /chromosome="2"  
 /clone="B121A12"

BASE COUNT 37461 a 30979 c 31992 g 37080 t 103 others  
 ORIGIN

Query Match 67.7%; Score 21; DB 2; Length 137615;  
 Best Local Similarity 82.8%; Pred. No. 32;  
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 CGGATTCACAGAGGTGATTTGCC 30  
 111 1 1111111 111 1111111  
 DB 23546 CGGACACTCATGTGGTTCATTGCC 23574

## RESULT 15

AC112349

LOCUS AC112349 155083 bp DNA linear HTG 13-JUL-2002  
 DEFINITION Rattus norvegicus clone CH230-336K24, \*\*\* SEQUENCING IN PROGRESS

AC112349

AC112349.2 GI:21737633  
 HTG: HTGS\_PHASE1.

KEYWORDS  
 SOURCE  
 ORGANISM

## REFERENCE

1 (bases 1 to 155083)  
 Muzny,D.M., Adams,C., Adio-oduola,B., Ali-osman,F.R., Allen,C., Alstbrooks,S.L., Amaralunga,H.C., Are,U.R., Ayele,M., Banks,T., Barbieri,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowe,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dihn,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Barnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Hollway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtenberg,O., Lieu,C., Liu,J., Liu,W., Lousleged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Messy,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Monabadi,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,S., Nguyen,N., Nickerson,E., Nwokenwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,

## REFERENCE

Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojupokan,I., Rolfe,M., Ruiz,S., Severy,G., Scherer,S., Scott,G., Shen,H., Shoochari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalton,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G., and Gibbs,R.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 155083)  
 Worley,K.C.  
 Direct Submission  
 Submitted (21-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 155083)  
 Worley,K.C.  
 Direct Submission  
 Submitted (13-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 On Jul 12, 2002 this sequence version replaced gi:18846589.  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: http://www.hgsc.bcm.tmc.edu/  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GPM1  
 Center clone name: CH230-336K24  
 ----- Summary Statistics  
 Sequencing vector: Plasmid;  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 115979 bases at least Q40  
 Consensus quality: 120179 bases at least Q30  
 Consensus quality: 123722 bases at least Q20  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft\_data.html).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 54 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 1  
 1065: contig of 1065 bp in length  
 1066: gap of unknown length  
 1166: contig of 1519 bp in length  
 2685: gap of unknown length  
 2785: contig of 1424 bp in length  
 4209: gap of unknown length  
 4309: contig of 1321 bp in length  
 5630: gap of unknown length  
 5729: contig of 1749 bp in length  
 7478: gap of unknown length  
 7479: contig of 1200 bp in length  
 7579: gap of unknown length  
 8779: contig of 1529 bp in length  
 8879: gap of unknown length  
 10408: contig of 1304 bp in length  
 10508: gap of unknown length  
 11812: contig of 1276 bp in length  
 11912: gap of unknown length  
 13187: contig of 1195 bp in length  
 13288: gap of unknown length  
 13288: contig of 1195 bp in length  
 14483: gap of unknown length  
 14583: contig of 1146 bp in length

\* 15729 15828: gap of unknown length  
\* 15829 16849: contig of 1021 bp in length  
\* 16850 16949: gap of unknown length  
\* 16950 18972: contig of 2023 bp in length  
\* 18973 19072: gap of unknown length  
\* 19073 20725: contig of 1653 bp in length  
\* 20726 20825: gap of unknown length  
\* 20826 22038: contig of 1213 bp in length  
\* 22039 22138: gap of unknown length  
\* 22139 24186: contig of 2048 bp in length  
\* 24187 24286: gap of unknown length  
\* 24287 25566: contig of 1280 bp in length  
\* 25567 25667: gap of unknown length  
\* 25667 27089: contig of 1423 bp in length  
\* 27090 27189: gap of unknown length  
\* 27190 29512: contig of 2323 bp in length  
\* 29513 29612: gap of unknown length  
\* 29613 32625: contig of 3013 bp in length  
\* 32626 32725: gap of unknown length  
\* 32726 34261: contig of 1536 bp in length  
\* 34262 34361: gap of unknown length  
\* 34362 36173: contig of 1812 bp in length  
\* 36174 36273: gap of unknown length  
\* 36274 38253: contig of 1980 bp in length  
\* 38254 38353: gap of unknown length  
\* 38354 39956: contig of 1603 bp in length  
\* 39957 40056: gap of unknown length  
\* 40057 41003: contig of 1547 bp in length  
\* 41004 41703: gap of unknown length  
\* 41704 43663: contig of 1960 bp in length  
\* 43664 43763: gap of unknown length  
\* 43764 45801: contig of 2038 bp in length  
\* 45802 45901: gap of unknown length  
\* 45902 48585: contig of 2684 bp in length  
\* 48586 48685: gap of unknown length  
\* 48686 50897: contig of 2212 bp in length  
\* 50898 50997: gap of unknown length  
\* 50999 53612: contig of 2615 bp in length  
\* 53613 53712: gap of unknown length  
\* 53713 55684: contig of 1972 bp in length  
\* 55685 55784: gap of unknown length  
\* 55785 58085: contig of 2301 bp in length  
\* 58086 58185: gap of unknown length  
\* 58186 61532: contig of 3347 bp in length  
\* 61533 61632: gap of unknown length  
\* 61633 63874: contig of 2242 bp in length  
\* 63875 63974: gap of unknown length  
\* 63975 67252: contig of 3278 bp in length  
\* 67253 67352: gap of unknown length  
\* 67353 69731: contig of 2379 bp in length  
\* 69732 69831: gap of unknown length  
\* 69832 72178: contig of 2347 bp in length  
\* 72179 72278: gap of unknown length  
\* 72279 74962: contig of 2684 bp in length  
\* 74963 75062: gap of unknown length  
\* 75063 78248: contig of 3186 bp in length  
\* 78249 78348: gap of unknown length  
\* 78349 81862: contig of 3514 bp in length  
\* 81863 81962: gap of unknown length  
\* 81963 86028: contig of 4066 bp in length  
\* 86029 86128: gap of unknown length  
\* 86129 88903: contig of 3775 bp in length  
\* 88904 90003: gap of unknown length  
\* 90004 93625: contig of 3622 bp in length  
\* 93626 93725: gap of unknown length  
\* 93726 97620: contig of 3895 bp in length  
\* 97621 97720: gap of unknown length  
\* 97721 102204: contig of 4484 bp in length  
\* 102205 102304: gap of unknown length  
\* 102305 107646: contig of 5342 bp in length  
\* 107647 107746: gap of unknown length  
\* 107747 112372: contig of 4626 bp in length  
\* 112373 112472: gap of unknown length

\* 112473 117432: contig of 4960 bp in length  
\* 117433 117532: gap of unknown length  
\* 117533 121402: contig of 3870 bp in length  
\* 121403 121502: gap of unknown length  
\* 121503 127185: contig of 5683 bp in length  
\* 127186 127285: gap of unknown length  
\* 127286 132130: contig of 4845 bp in length  
\* 132131 132230: gap of unknown length  
\* 132231 140048: contig of 7818 bp in length  
\* 140049 140148: gap of unknown length  
\* 140149 146515: contig of 6367 bp in length  
\* 146516 146615: gap of unknown length  
\* 146616 155083: contig of 8468 bp in length.

Query Match 66.5%; Score 20.6; DB 2; Length 155083;  
Best Local Similarity 85.2%; Pred. No. 52;  
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 3 GGATTCTCACATGAGGGTCATTTC 29  
DB 4763 GGATTGGGACATGGGGTCATTTC 4789

Search completed: March 23, 2003, 16:04:28  
Job time : 339.909 secs

GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 23, 2003, 13:56:37 ; Search time 36.0455 Seconds  
(without alignments)  
1936.775 Million cell updates/sec

Title: US-09-873-546-3  
Perfect score: 31  
Sequence: 1 gcggaattccacatgaggtgcatggccc 31

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues  
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_101002:\*

1: /SID52/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:\*

2: /SID52/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:\*

3: /SID52/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:\*

4: /SID52/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:\*

5: /SID52/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:\*

6: /SID52/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:\*

7: /SID52/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:\*

8: /SID52/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:\*

9: /SID52/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:\*

10: /SID52/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:\*

11: /SID52/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:\*

12: /SID52/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:\*

13: /SID52/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:\*

14: /SID52/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:\*

15: /SID52/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:\*

16: /SID52/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:\*

17: /SID52/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:\*

18: /SID52/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:\*

19: /SID52/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:\*

20: /SID52/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:\*

21: /SID52/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:\*

22: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:\*

23: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:\*

24: /SID52/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	74.2	1249	24	ABO72525	Human MDDT encodin
2	74.2	3061	24	ABO72648	Human MDDT encodin
3	20	64.5	21521	22 ABA21357	Human nervous syst
4	20	64.5	21521	22 AAK78175	Human immune/haema
5	19.8	63.9	294	24 ABN20378	Human OREF polynuc
6	19.8	63.9	433	22 AAK52951	Human polynucleoti
7	19.8	63.9	527	22 AAK51967	Human polynucleoti
8	19.8	63.9	710	22 AAF58344	Human GTP-binding
9	19.8	63.9	2484	22 ABA20286	Human nervous syst

C 10	19.8	63.9	2484	22 ABA20287	Human nervous syst
C 11	19.4	62.6	3098	20 AAZ00362	Nucleotide sequenc
C 12	19.4	62.6	3098	22 AAF89024	Rat FAP1 coding s
C 13	19.2	61.9	340	24 ABN95508	Gene #2006 used to
C 14	19.2	61.9	340	24 ABK64471	Human benign prost
C 15	19.2	61.3	12932	23 ABK42245	Genomic sequence #
C 16	18.6	60.0	24259	22 AAS46692	Tumour suppressor
C 17	18.4	59.4	787	20 AAZ17527	Human gene express
C 18	18.4	59.4	2824	22 AAH18429	Human cDNA sequenc
C 19	18.4	59.4	2936	22 AAS08646	Human cDNA encodin
C 20	18.4	59.4	3098	21 AAA23459	CDNA encoding huma
C 21	18.4	59.4	3844	22 AAH76202	Human drug metabol
C 22	18.4	59.4	3910	22 AAS15815	Human cDNA encodin
C 23	18.4	59.4	3912	22 AAH14598	Human cDNA sequenc
C 24	18.4	59.4	6663	19 AAH27145	Nucleotide sequenc
C 25	18.4	59.4	6663	22 AAD04197	Murine haemopoieti
C 26	18.4	59.4	11832	19 AAH27148	Nucleotide sequenc
C 27	18.4	59.4	11832	22 AAD04198	Murine NR6 genomic
C 28	18.2	58.7	335	22 AAF66850	Novel human polynu
C 29	18.2	58.7	470	24 ABN96653	Gene #3151 used to
C 30	18.2	58.7	470	24 ABK64694	Human benign prost
C 31	18.2	58.7	470	24 ABK68241	Kidney cancer rela
C 32	18.2	58.7	470	24 ABK68479	Kidney cancer rela
C 33	18.2	58.7	470	24 ABK69695	Prostate cancer re
C 34	18.2	58.7	661	24 ABK65887	Arabidopsis thalia
C 35	18.2	58.7	1325	21 AAC32810	Arabidopsis thalia
C 36	18.2	58.7	1809	21 AAC59514	Human secreted pro
C 37	18.2	58.7	9817	24 ABK33369	Human immune syste
C 38	18	58.1	353	14 AAO60348	Human brain Expres
C 39	18	58.1	7653	24 AAD22153	PUC19-Sep1.5 plas
C 40	18	58.1	7824	17 AAT33872	Human RAP1 cDNA.
C 41	18	58.1	7943	19 AAV28518	FRAP (tor1) cDNA.
C 42	17.8	57.4	296	20 AAV88071	ESF clone FY127.
C 43	17.8	57.4	514	22 ABA51413	Human foetal liver
C 44	17.8	57.4	514	22 AAK09711	Human brain expres
C 45	17.8	57.4	514	22 AAK35605	Human bone marrow

## ALIGNMENTS

RESULT 1	ABO72525/c	ABO72525 standard; cDNA: 1249 BP.
ID	ABO72525;	
AC	ABO72525;	
XX		
XX		
DT	03-SEP-2002 (first entry)	
DE		
XX	Human MDDT encoding cDNA SEQ ID NO 77.	
KW	Human: MDDT; disease detection and treatment molecule polynucleotide;	
KW	proliferative disorder; hepatitis; psoriasis; cancer; AIDS;	
KW	autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;	
KW	rheumatoid arthritis; transgenic; gene therapy; antileukosclerotic;	
KW	hepatotropic; antiinflammatory; antipsoriatic; cytosolic; anti-HIV;	
KW	antiallergic; antianemic; antisthmatic; antiatherosclerotic; antignout;	
KW	neuroprotective; antineumatic; antiarthritic; gene; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	WO200240715-A2.	
XX		
PD	23-MAY-2002.	
XX		
XX	06-SEP-2001; 2001WO-US27628.	
PF		
XX		
XX	06-SEP-2000; 2000US-230505P.	
PR	06-SEP-2000; 2000US-230514P.	
PR	06-SEP-2000; 2000US-230515P.	
PR	06-SEP-2000; 2000US-230517P.	
PR	06-SEP-2000; 2000US-230518P.	
PR	06-SEP-2000; 2000US-230519P.	

PR 06-SEP-2000: 2000US-230595P.  
PR 06-SEP-2000: 2000US-230597P.  
PR 06-SEP-2000: 2000US-230598P.  
PR 06-SEP-2000: 2000US-230599P.  
PR 06-SEP-2000: 2000US-230610P.  
PR 06-SEP-2000: 2000US-230655P.  
PR 06-SEP-2000: 2000US-230988P.  
PR 06-SEP-2000: 2000US-230989P.  
PR 07-SEP-2000: 2000US-230951P.  
PR 07-SEP-2000: 2000US-231163P.  
PR 07-SEP-2000: 2000US-231167P.  
XX  
PA (INCYTE GENOMICS INC.  
XX  
PI Jackson S, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL;  
PI Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE, Dahl CR;  
PI Momiyama MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM;  
PI Gerstlin EH, Peralta CH, David MH, Panzer SR, Flores V, Dafio A;  
PI Marwaha R, Chen AJ, Chang SC, Au AP, Inman RR;  
XX  
XX MPI: 2002-527544/56.  
XX P-PSDB: ABP51307.

Novel human disease detection and treatment polypeptide, useful in diagnosis, prevention or treatment of cell proliferative disorders e.g. arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder e.g. AIDS

Claim 1: Page 350-351: 618pp; English.

The invention relates to an isolated human disease detection and treatment (MDPT) polypeptide (I) selected from a polypeptide having a sequence selected from 254 sequences (ABP51231-ABP51484) given in the specification, a naturally occurring polypeptide comprising a sequence having at least 90% identity to (I) or a biologically active or immunogenic fragment of (I). (I) is useful for screening a compound for effectiveness as an agonist or antagonist, for screening a compound that specifically binds (I) or modulates the activity of (I), and for preparing a polyclonal or monoclonal antibody by hybridoma technology. Nucleic acids (II) (ABQ72449-ABQ72700) encoding (I) are useful for screening a compound for effectiveness in altering expression of a target polynucleotide comprising. Oligonucleotides and antibodies are useful for detecting MDDT in a sample or for assessing toxicity of a test compound, in a diagnostic test for a condition or a disease associated with the expression of MDDT in a biological sample, for detecting (I) in a sample, and for purifying (I) from a sample. A composition comprising (I), an agonist or antagonist is useful for treating a disease or condition associated with decreased or increased expression of functional MDDT. (I) or (II) are useful for diagnosing, treating or preventing disorders associated with aberrant expression of MDDT, where the disorders are selected from a cell proliferative disorder such as arteriosclerosis, cirrhosis, hepatitis, psoriasis, and cancer and an autoimmune/inflammatory disorder such as AIDS, Addison's disease, allergy, anaemia, asthma, atherosclerosis, gout, multiple sclerosis or rheumatoid arthritis. (II) are useful for creating knockin humanised animals or transgenic animals to model human diseases, in somatic or germ-line gene therapy, to generate a transcript image of a tissue or cell type, for detecting differences in the chromosomal location due to translocation or inversion among normal, carrier or affected individuals and as hybridisation probes for mapping naturally occurring genomic sequences.

Sequence 1249 BP: 230 A: 436 C: 355 G: 228 T: 0 other:

Query Match 74.2%; Score 23; DB 24; Length 1249;  
Best Local Similarity 100.0%; Pred. No. 0.66;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 CTCACATGAGGTCATTGCCC 31  
DB 647 CTCACATGAGGTCATTGCCC 625

RESULT 2  
ABQ72648/c  
ID ABQ72648 standard; cDNA: 3061 BP.  
XX  
XX  
AC ABQ72648;  
XX  
DT 03-SEP-2002 (first entry)  
XX  
XX  
DE Human MDDT encoding cDNA SEQ ID NO 200.  
XX  
XX  
KW Human: MDDT: disease detection and treatment molecule polynucleotide;  
KW proliferative disorder; hepatitis; psoriasis; cancer; AIDS;  
KW autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;  
KW rheumatoid arthritis; transgenic; gene therapy; antiarteriosclerosis;  
KW hepatotropic; antiinflammatory; antipsoriatic; cytostatic; anti-HIV;  
KW antiallergic; antianaemic; antiasthmatic; antiatherosclerotic; antiout;  
KW neuroprotective; antirheumatic; antiarthritic; gene; ss.  
XX  
OS Homo sapiens.  
XX  
XX WO2002040715-A2.  
XX  
XX  
PD 23-MAY-2002.  
XX  
XX  
PF 06-SEP-2001: 2001WO-US27628.  
XX  
XX  
PR 06-SEP-2000: 2000US-230505P.  
PR 06-SEP-2000: 2000US-230514P.  
PR 06-SEP-2000: 2000US-230515P.  
PR 06-SEP-2000: 2000US-230517P.  
PR 06-SEP-2000: 2000US-230518P.  
PR 06-SEP-2000: 2000US-230519P.  
PR 06-SEP-2000: 2000US-230595P.  
PR 06-SEP-2000: 2000US-230597P.  
PR 06-SEP-2000: 2000US-230598P.  
PR 06-SEP-2000: 2000US-230599P.  
PR 06-SEP-2000: 2000US-230610P.  
PR 06-SEP-2000: 2000US-230655P.  
PR 06-SEP-2000: 2000US-230988P.  
PR 06-SEP-2000: 2000US-230989P.  
PR 06-SEP-2000: 2000US-230951P.  
PR 06-SEP-2000: 2000US-231163P.  
PR 07-SEP-2000: 2000US-231167P.

(INCYTE GENOMICS INC.

Jackson S, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL;  
PI Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE, Dahl CR;  
PI Momiyama MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM;  
PI Gerstlin EH, Peralta CH, David MH, Panzer SR, Flores V, Dafio A;  
PI Marwaha R, Chen AJ, Chang SC, Au AP, Inman RR;  
XX  
XX MPI: 2002-527544/56.  
XX P-PSDB: ABP51431.

Novel human disease detection and treatment polypeptide, useful in diagnosis, prevention or treatment of cell proliferative disorders e.g. arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder e.g. AIDS

Claim 1: Page 414: 618pp; English.

The invention relates to an isolated human disease detection and treatment (MDPT) polypeptide (I) selected from a polypeptide having a sequence selected from 254 sequences (ABP51231-ABP51484) given in the specification, a naturally occurring polypeptide comprising a sequence having at least 90% identity to (I) or a biologically active or immunogenic fragment of (I). (I) is useful for screening a compound for effectiveness as an agonist or antagonist, for screening a compound that specifically binds (I) or modulates the activity of (I), and for preparing a polyclonal or monoclonal antibody by hybridoma technology. Nucleic acids (II) (ABQ72449-ABQ72700) encoding (I) are useful for screening a compound for effectiveness in altering expression of a target

CC polynucleotide comprising. Oligonucleotides and antibodies are useful for  
CC detecting MDPF in a sample or for assessing toxicity of a test compound,  
CC in a diagnostic test for a condition or a disease associated with the  
CC expression of MDPF in a biological sample, for detecting (I) in a sample,  
CC and for purifying (I) from a sample. A composition comprising (I), an  
CC agonist or antagonist is useful for treating a disease or condition  
CC associated with decreased or increased expression of functional MDPF.  
CC (I) or (II) are useful for diagnosing, treating or preventing disorders  
CC associated with aberrant expression of MDPF, where the disorders are  
CC selected from a cell proliferative disorder such as arteriosclerosis,  
CC cirrhosis, hepatitis, psoriasis, and cancer and an  
CC autoimmune/inflammatory disorder such as AIDS, Addison's disease,  
CC allergy, anaemia, asthma, atherosclerosis, gout, multiple sclerosis or  
CC rheumatoid arthritis. (II) are useful for creating knockin humanised  
CC animals or transgenic animals to model human diseases, in somatic or  
CC germ-line gene therapy, to generate a transcript image of a tissue or cell  
CC type, for detecting differences in the chromosomal location due to  
CC translocation or inversion among normal, carrier or affected individuals  
CC and as hybridisation probes for mapping naturally occurring genomic  
CC sequences.

XX Sequence 3061 BP; 606 A; 840 C; 927 G; 688 T; 0 other;

Query Match 74.2%; Score 23; DB 24; Length 3061;  
Best Local Similarity 100.0%; Pred. No. 0.79;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 9 CTCACATGAGGTCGATTGCC 31  
|||||

Db 649 CTCACATGAGGTCGATTGCC 627

### RESULT 3

ABA21357/C  
ID ABA21357 standard; DNA; 21521 BP.

XX ABA21357;

DT 23-JAN-2002 (first entry)

XX Human nervous system related polynucleotide SEQ ID NO 13688.

KW Human; neurotropic; neuroprotective; cytostatic; dermatological; virocidic;  
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;  
KW antiparkinsonian; antischlicking; antianaemic; antiarthritic; cancer;  
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;  
KW antiallergic; antidiabetic; antitumor; anticonvulsant; antitumoral;  
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.

OS Homo sapiens.

XX WO200159063-A2.

XX 16-AUG-2001.

PE 17-JAN-2001; 2001WO-US01334.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 14-AUG-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231245.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235835.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 20-OCT-2000; 2000US-0242221.  
PR 01-NOV-2000; 2000US-0244617.

PR	08-NOV-2000;	2000US-0246474.
PR	08-NOV-2000;	2000US-0246475.
PR	08-NOV-2000;	2000US-0246476.
PR	08-NOV-2000;	2000US-0246477.
PR	08-NOV-2000;	2000US-0246478.
PR	08-NOV-2000;	2000US-0246523.
PR	08-NOV-2000;	2000US-0246524.
PR	08-NOV-2000;	2000US-0246525.
PR	08-NOV-2000;	2000US-0246526.
PR	08-NOV-2000;	2000US-0246527.
PR	08-NOV-2000;	2000US-0246528.
PR	08-NOV-2000;	2000US-0246532.
PR	08-NOV-2000;	2000US-0246609.
PR	08-NOV-2000;	2000US-0246610.
PR	08-NOV-2000;	2000US-0246611.
PR	08-NOV-2000;	2000US-0246613.
PR	17-NOV-2000;	2000US-0249207.
PR	17-NOV-2000;	2000US-0249208.
PR	17-NOV-2000;	2000US-0249209.
PR	17-NOV-2000;	2000US-0249210.
PR	17-NOV-2000;	2000US-0249211.
PR	17-NOV-2000;	2000US-0249212.
PR	17-NOV-2000;	2000US-0249213.
PR	17-NOV-2000;	2000US-0249214.
PR	17-NOV-2000;	2000US-0249215.
PR	17-NOV-2000;	2000US-0249216.
PR	17-NOV-2000;	2000US-0249217.
PR	17-NOV-2000;	2000US-0249218.
PR	17-NOV-2000;	2000US-0249219.
PR	17-NOV-2000;	2000US-0249244.
PR	17-NOV-2000;	2000US-0249245.
PR	17-NOV-2000;	2000US-0249246.
PR	17-NOV-2000;	2000US-0249265.
PR	17-NOV-2000;	2000US-0249267.
PR	17-NOV-2000;	2000US-0249297.
PR	17-NOV-2000;	2000US-0249299.
PR	17-NOV-2000;	2000US-0249300.
PR	01-DEC-2000;	2000US-0250391.
PR	01-DEC-2000;	2000US-0251160.
PR	03-DEC-2000;	2000US-0251030.
PR	05-DEC-2000;	2000US-0251088.
PR	05-DEC-2000;	2000US-0251679.
PR	06-DEC-2000;	2000US-0251479.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251989.
PR	08-DEC-2000;	2000US-0251990.
PR	11-DEC-2000;	2000US-0254097.
PR	05-JAN-2001;	2001US-0259678.
PR	(HUMA-) HUMAN GENOME SCI INC.	
XX	Rosen CA, Barash SC, Ruben SM;	
XX	WPI; 2001-541565/60.	
XX		
XX	Nucleic acids encoding 3224 human nervous system antigen polypeptides,	
PT	useful for preventing, diagnosing and/or treating nervous system	
PT	cancers and metastases -	
XX		
PS	Disclosure; SEQ ID NO 13688; 1701pp + Sequence Listing; English.	
XX		
CC	The invention relates to novel genes (ABA11004-ABA21534) and proteins	
CC	(AB14678-AB181001) useful for preventing, treating or ameliorating	
CC	medical conditions e.g. by protein or gene therapy. The genes are	
CC	isolated from a range of human tissues disclosed in the specification.	
CC	The nucleic acids, proteins, antibodies and (ant)agonists are useful	
CC	in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast	
CC	and ovarian cancer and other cancers of the adrenal gland, bone, bone	
CC	marrow, breast, gastrointestinal tract, liver, lung, or urogenital;	
CC	(b) immune disorders e.g. Addison's disease, allergies, autoimmune	
CC	hemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's	
CC	disease, multiple sclerosis, rheumatoid arthritis and ulcerative	
CC	colitis; (c) cardiovascular disorders such as myocardial ischaemias;	

[illegible]



```

PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234297.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.

PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI; 2001-483426/52.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
useful for preventing, diagnosing and/or treating cancers and
metastasis -
XX
PS Disclosure; SEQ ID NO 32987; 3071pp + Sequence listing; English.
XX
XX
CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 21521 BP; 6512 A; 4334 C; 4422 G; 6253 T; 0 other;

Query Match 64.5%; Score 20; DB 22; Length 21521;
Best Local Similarity 82.1%; Pred No. 30;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCGGAATTCACATGAGGCGTCATTG 28
Db 16126 GCGGAATTCACATGAGGAGCAGTTG 16099

RESULT 5
ABN20378/c

```

ABN20378 standard; cDNA: 294 BP.  
ABN20378;  
24-JUN-2002 (first entry)  
Human ORFX polynucleotide sequence SEQ ID NO:9233.  
Human: open reading frame: ORFX; gene therapy: cancer: cirrhosis; hyperproliferative disorder: psoriasis; benign tumour: haemorrhage; degenerative disorder; osteoarthritis; neurodegenerative disorder; cardiovascular disease; diabetes mellitus; systemic lupus erythematosus; hypertension; hypothyroidism; cholesterol ester storage disease; immune deficiency; immune disorder; infectious disease; autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis; myasthenia gravis; gene: ss.  
Homo sapiens.  
MO200192523-A2.  
06-DEC-2001.  
29-MAY-2001; 2001WO-US10836.  
30-MAY-2000; 2000US-206132P.  
29-AUG-2000; 2000US-228716P.  
(CURA-) CURAGEN CORP.  
Shimkets RA, Leach MD;  
WPI: 2002-106308/14.  
P-PSDB: ABP04626.  
Novel human polypeptides and polynucleotides useful for diagnosing, preventing and treating cardiovascular disease, neurodegenerative, hyperproliferative disorders and autoimmune disorders  
Disclosure: SEQ ID 9233; 1037Pp: English.  
The present invention describes substantially purified human proteins (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1 in the specification). ABN15762 to ABN27252 encode the human ORFX proteins given in ABP00010 to ABP11500. ORFX proteins are useful for treating or preventing a pathology associated with an ORFX-associated disorder in humans, and in the manufacture of a medicament for treating a syndrome associated with ORFX-associated disorder. ORFX polynucleotide sequences can be used in gene therapy. ORFX sequences can be used in the treatment of cancer, hyperproliferative disorders, cirrhosis of liver, psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage, osteoarthritis, neurodegenerative disorders, diabetes mellitus, systemic lupus erythematosus, hypertension, hypothyroidism, cholesterol ester storage disease, various immune deficiencies and disorders, infectious diseases, autoimmune disorders such as multiple sclerosis, rheumatoid arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. ORFX proteins are also useful for treating burns, incisions, ulcers, for treating osteoporosis, bone degenerative disorders, or periodontal disease, and for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues and conditions resulting from systemic cytokine damage.  
N.B. The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
Sequence 294 BP; 59 A; 83 C; 84 G; 67 T; 1 other;

OY 1 GCGGAATTCACATGAGGGTGCATTGTGCC 31  
II IIIIIIIIIII I I IIII IIIII I  
Db 225 GCCGAATTCACCAGGTGGCTGCTTTGGCC 195

RESULT 6  
AAK52951/c  
ID AAK52951 standard; cDNA: 433 BP.  
XX AAK52951;  
DT 06-NOV-2001 (first entry)  
DE Human polynucleotide SEQ ID NO 2480.  
XX  
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
XX nervous system disorder; arthritis; inflammation; ss.  
OS Homo sapiens.  
PN WO200157190-A2.  
PD 09-AUG-2001.  
XX  
PF 05-FEB-2001; 2001WO-US04098.  
PR 03-FEB-2000; 2000US-0496914.  
PR 27-APR-2000; 2000US-0560875.  
PR 20-JUN-2000; 2000US-0598075.  
PR 19-JUL-2000; 2000US-0620325.  
PR 01-SEP-2000; 2000US-0654936.  
PR 15-SEP-2000; 2000US-0663561.  
PR 20-OCT-2000; 2000US-0693325.  
PR 30-NOV-2000; 2000US-0728422.  
PA (HYSE-) HYSEQ INC.

P1 Tang YT, Liu C, Dermanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
P2 Zhao QH, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
P1 Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
XX  
DR WPI: 2001-476283/51.  
P-PSDB: AAM79818.

XX Nucleic acids encoding polypeptides with cytokine-like activities,  
PT useful in diagnosis and gene therapy -  
PT Claim 1; Page 4758-4759; 6221pp; English.

XX The invention relates to polynucleotides (AAK51456-AAK53435) and the  
CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC active/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: Records for SRO ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
CC (AAM80020) are omitted as the relevant pages from the sequence listing  
CC were missing at the time of publication.

```
Query Match      63.9%; Score 19.8; DB 24; Length 294;
Best Local Similarity 77.4%; Pred. No. 16;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
```

Best Local Similarity 91.3%; Pred. NO. 17;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0.





PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
DR WPI; 2001-541565/60.  
XX  
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating nervous system  
PT cancers and metastases -  
XX  
PS Disclosure; SEQ ID NO 12617; 1701bp + Sequence Listing; English.  
XX  
CC The invention relates to novel genes (ABAI1004-ABA21534) and proteins  
CC (ABAI1678-ABAI18001) useful for preventing, treating or ameliorating  
CC medical conditions e.g., by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
CC hemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 2484 BP; 440 A; 906 C; 734 G; 404 T; 0 other;

Query Match 63.9%; Score 19.8; DB 22; Length 2484;  
Best Local Similarity 91.3%; Pred. No. 24;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 9 CTCACATGAGGATGATTTGCC 31  
DB 2036 CTCACAGGATGATTTGCC 2014

RESULT 10  
ABA20287/c  
ID ABA20287 standard; DNA: 2484 BP.  
XX  
AC ABA20287;  
XX  
DT 23-JAN-2002 (first entry)  
XX  
DE Human nervous system related polynucleotide SEQ ID NO 12618.  
XX  
KW Human; nootropic; neuroprotective; cytosolic; dermatological; virocidic;  
KW immunosuppressive; anti-infective; anti-HIV; antibacterial; vulnery;  
KW antiparkinsonian; antischizoid; antianemic; antitubercular; cancer;  
KW antineoplastic; hepatotropic; neuroprotective; antiinflammatory;  
KW antiallergic; antidiabetic; antileukemic; anticonvulsant; antifungal;  
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200159063-A2.  
XX  
PD 16-AUG-2001.

XX  
PF 17-JAN-2001; 2001WO-US01334.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205513.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220963.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.

PT	Nucleic acids encoding 3224 human nervous system antigen polypeptides, useful for preventing, diagnosing and/or treating nervous system cancers and metastases -	PT
XX	Disclosure; SEQ ID NO 12618; 1701bp + Sequence Listing; English.	XX
XX	The invention relates to novel genes (ABA11004-ABA21534) and proteins (ABA14678-ABA18001) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification.	XX
CC	The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital;	CC
CC	(b) immune disorders e.g. Addison's disease, allergies, autoimmune disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections.	CC
CC	Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at <a href="http://wipo.int/pub/published_pctl_sequences">ftp://wipo.int/pub/published_pctl_sequences</a> .	CC
XX	Sequence 2484 BP; 440 A; 906 C; 734 G; 404 T; 0 other;	XX
XX	Query Match 63.9%; Score 19.8; DB 22; Length 2484; Best local Similarity 91.3%; Pred. No. 24;	XX
XX	Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;	XX
QY	9 CTCACATGAGGTCATTTGCC 31	QY
DB	2036 CTCACAGATGTCATTTGCC 2014	DB
AAZ00362/c	AAZ00362 standard; DNA; 3098 BP.	AAZ00362/c
AAZ00362;		AAZ00362;
XX	26-OCR-1999 (first entry)	XX
DE	Nucleotide sequence of rat rnfatp1.	DE
KW	Fatty acid transport protein; FATP; long chain fatty acid; LCFA; fatty acid; FATP biosynthesis; obesity; diabetes; heart disease; ss.	KW
OS	Rattus norvegicus.	OS
XX		XX
PN	W09936537-A2.	PN
XX		XX
PD	22-JUL-1999.	PD
PE	14-JAN-1999; 99WO-US00182.	PE
XX		XX
PR	14-JAN-1999; 99US-0232201.	PR
PR	15-JAN-1998; 98US-0071374.	PR
PR	20-JUL-1998; 98US-0093491.	PR
PR	04-DEC-1998; 98US-0110941.	PR
PR	14-JAN-1999; 99US-0232195.	PR
PR	14-JAN-1999; 99US-0232197.	PR
PR	14-JAN-1999; 99US-0232200.	PR
XX		XX
PA	(MILL-) MILLENNIUM PHARM INC.	PA
PA	(WHED) WHITEHEAD INST BIOMEDICAL RES.	PA
XX		XX
PI	Giмено RE, Hirsch DJ, Lodish HF, Stahl A, Tartaglia LA; WPI, 1999-444398/37.	PI
XX		XX
DR	P-PSDB; AAY14952.	DR
XX		XX

PT Fatty acid transport proteins and related polynucleotides, useful  
PT for treating obesity, diabetes and heart disease  
XX  
XX  
PS Disclosure; Fig 56; 255pp; English.  
XX  
CC The invention provides a family of fatty acid transport proteins (FATPs)  
CC that mediate transport of long chain fatty acids (LCFAs) across cell  
CC membranes into cells. Human and murine FATP proteins and nucleic acids  
CC encoding the proteins are provided. The FATP proteins can be produced  
CC by standard recombinant methodology. Fatty acid uptake by cells can be  
CC modulated by modulating biosynthesis of FATP proteins especially FATP6.  
CC In particular, antisense oligonucleotides can be used to modulate FATP  
CC biosynthesis. Modulation of FATP6 is useful for inhibiting fatty acid  
CC uptake in cardiac muscle of humans. Agents can be directed to cardiac  
CC muscle or liver by administration of a complex of the agent and a FATP6  
CC binding moiety. DNA encoding FATP proteins can be used as a reference  
CC used in detecting variant alleles or homologues. Altering the LCFA uptake  
CC by administering an inhibitor or enhancer of FATP transport function in  
CC the small intestine can decrease or increase calories available as fats,  
CC and can decrease or increase circulating fatty acids. Blocking the  
CC function of FATP4 and also FATP2, is useful for treating obesity,  
CC diabetes and heart disease.  
XX  
SQ Sequence 3098 BP; 614 A; 911 C; 903 G; 670 T; 0 other;  
XX  
Query Match 62.6%; Score 19.4; DB 20; Length 3098;  
Best Local Similarity 95.2%; Pred. No. 39;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
XX  
OY 9 CTCACATGAGGTCATTGTC 29  
DB 2433 CTGACATGAGGTCATTGTC 2413  
II |||||  
RESULT 12  
AA89024/C  
ID AA89024 standard; DNA: 3098 BP.  
XX  
XX AA89024;  
XX  
DT 06-JUL-2001 (first entry)  
XX  
DE Rat FATP1 coding sequence SEQ ID NO: 58.  
XX  
XX Fatty acid transport protein; FATP; human; mouse; rat; rice blast fungus;  
KM yeast; fat absorption; obesity; diabetes; heart disease; hyperlipidaemia;  
KM weight control; tuberculosis; TB; anti-fungal; ds.  
XX  
XX Rattus norvegicus.  
XX  
XX WO200121795-A2.  
XX  
XX 29-MAR-2001.  
XX  
XX 21-SEP-2000; 2000WO-US25891.  
XX  
XX 23-SEP-1999; 99US-0405504.  
PR 23-SEP-1999; 99US-0405505.  
PR 16-DEC-1999; 99US-0465280.  
PR 17-FEB-2000; 2000US-0506252.  
PR 06-JUL-2000; 2000US-0611197.  
XX  
XX (WHED ) WHITEHEAD INST BIOMEDICAL RES.  
PA (MILL-) MILLENNIUM PHARM INC.  
XX  
XX Stahl A, Hirsch DJ, Lodish HF, Gimeno RE, Tartaglia LA;  
PI  
XX WPI; 2001-354783/37.  
DR P-PSDB; AA883252.  
XX  
XX New fatty acid transport proteins (FATPs) useful for the manufacture of  
PT medicament for treating obesity, diabetes and heart disease -  
XX

PS Disclosure; Fig 56; 287pp; English.  
XX  
XX The present invention provides the protein and coding sequences of fatty  
CC acid transport proteins (FATPs) from a number of species, including  
CC FATP1, FATP2, FATP3, FATP4 and FATP6 from the human, FATP1-FATP5  
CC from the mouse, FATP4 and b from C. elegans, and FATP from Aspergillus  
CC nidulans, Dirosophila, zebrafish, Magnaporthe grisea, Mycobacterium  
CC tuberculosis and Cochliobolus heterostrophus. The FATP from M.  
CC tuberculosis can be used to identify inhibitors which can then be used to  
CC treat TB. That from M. grisea (also known as rice blast fungus) can be  
CC used to develop anti-fungal agents capable of preventing infection of  
CC rice. Those from the human can be used to develop treatments for  
CC diabetes, heart disease, obesity, hyperlipidaemia and weight control. The  
CC present sequence is one of the sequences described in the exemplification  
CC of the invention.  
XX  
SQ Sequence 3098 BP; 614 A; 911 C; 903 G; 670 T; 0 other;  
XX  
Query Match 62.6%; Score 19.4; DB 22; Length 3098;  
Best Local Similarity 95.2%; Pred. No. 39;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
XX  
OY 9 CTCACATGAGGTCATTGTC 29  
DB 2433 CTGACATGAGGTCATTGTC 2413  
II |||||  
RESULT 13  
ABN95508  
ID ABN95508 standard; DNA: 340 BP.  
XX  
XX ABN95508;  
XX  
DT 13-AUG-2002 (first entry)  
XX  
DE Gene #2006 used to diagnose liver cancer.  
XX  
XX Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;  
KM metastatic liver tumour; cytostatic; expression profile; disease state;  
KM disease progression; drug toxicity; drug efficacy; drug metabolism.  
XX  
XX Homo sapiens.  
XX  
XX WO200229103-A2.  
XX  
XX 11-APR-2002.  
XX  
XX 02-OCT-2001; 2001WO-US30589.  
XX  
XX 02-OCT-2000; 2000US-237054P.  
PR  
XX (GENE-) GENE LOGIC INC.  
XX  
XX Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;  
PI  
XX WPI; 2002-426119/45.  
XX  
XX Diagnosing and detecting the progression of liver cancer,  
PT hepatocellular carcinoma or metastatic liver tumor in a patient,  
PT involves detecting the level of expression of two or more genes in a  
PT liver tissue sample -  
XX  
XX Claim 1; SEQ ID NO 2006; 298pp; English.  
XX  
XX The invention relates to a novel method for diagnosing and detecting the  
CC progression of liver cancer, hepatocellular carcinoma or metastatic liver  
CC tumour in a patient, and differentiating metastatic liver cancer from  
CC hepatocellular carcinoma in a patient, involving detecting the level of  
CC expression of two or more genes represented in ABN93503-ABN97455 in a  
CC tissue sample. The method of the invention has hepatotropic, and  
CC cytostatic activity. The method is useful for diagnosing and detecting  
CC the progression of liver cancer, hepatocellular carcinoma and metastatic  
CC liver carcinoma in a patient. The method is useful for identifying

CC expression profiles which serve as useful diagnostic markers as well as  
CC markers that can be used to monitor disease states, disease progression,  
CC drug toxicity, drug efficacy and drug metabolism.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
CC XX

SO Sequence 340 BP; 84 A; 79 C; 98 G; 76 T; 3 other;

Query Match 61.9%; Score 19.2; DB 24; Length 340;  
Best Local Similarity 87.5%; Pred. No. 31;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 7 TTCTCATGAGGTCGTCATTGCC 30  
||||| 11 |||||||||||||||  
DB 238 TTCTTCAGAGGTCGTCATTGCC 261

RESULT 14  
ABK64471  
ABK64471 standard; DNA; 340 BP.

AC ABK64471;  
XX  
DT 18-JUN-2002 (first entry)  
XX  
DE Human benign prostatic hyperplasia gene #366.  
XX  
KW Human; benign prostatic hyperplasia; BPH; prostate cancer; gene; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200212440-A2.  
XX  
PD 14-FEB-2002.  
XX  
PE 07-AUG-2001; 2001WO-US24708.  
XX  
PF 07-AUG-2000; 2000US-223323P.  
XX  
PR 05-JUN-2001; 2001US-0873319.  
XX  
PA (GENE-) GENE LOGIC INC.  
XX (NISB) JAPAN TOBACCO INC.  
XX  
PI Munger WE, Kulkarni P, Getzenberg RH, Maga I, Yamamoto J;  
XX WPI; 2002-257476/30.  
XX  
XX Identifying drugs for and diagnosing benign prostatic hyperplasia, by  
XX detecting expression levels of one or more genes in prostate cells from  
XX patient that are differentially regulated compared to normal prostate  
XX cells -  
XX  
PS Disclosure; Page 207; 444pp; English.  
XX  
XX The invention relates to a method of diagnosing (I) the onset or  
XX progression of benign prostatic hyperplasia (BPH), or screening (II) for  
XX or identifying an agent that modulates the onset or progression of BPH.  
XX The method is based on changes in gene expression in BPH tissue isolated  
XX from patients exhibiting different clinical states of prostate  
XX hyperplasia as compared to normal prostate tissue. (I) comprises  
XX detecting the expression levels of one or more genes in prostate cells  
XX from the subject that are differentially regulated compared to normal  
XX prostate cells. (II) comprises preparing a first gene expression profile  
XX of BPH cells or BPH-like cell population, exposing the cells to the  
XX agent, preparing a second gene expression profile of the agent exposed  
XX cells, and comparing the first and second gene expression profiles  
XX (I) is useful for diagnosing the onset or progression of BPH. (II) is  
XX useful for identifying an agent that modulates the onset or progression  
XX of BPH. The methods are useful to present information identifying  
XX the expression level in a tissue or cells, by comparing the expression  
XX level of genes given in the specification in the tissue or cells to the  
XX level of expression of gene in the database, and displaying the

CC expression levels of at least one gene in the tissue or cell sample  
CC compared to the expression level in BPH. Agents using (II) are useful for  
CC treating BPH or prostate cancer. ABK64106-ABK64860 represent human  
CC benign prostatic hyperplasia gene sequences of the invention.  
CC XX

SO Sequence 340 BP; 84 A; 79 C; 98 G; 76 T; 3 other;

Query Match 61.9%; Score 19.2; DB 24; Length 340;  
Best Local Similarity 87.5%; Pred. No. 31;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 7 TTCTCATGAGGTCGTCATTGCC 30  
||||| 11 |||||||||||||||  
DB 238 TTCTTCAGAGGTCGTCATTGCC 261

RESULT 15  
ABK42245/C  
ID ABK42245 standard; DNA; 12932 BP.  
XX  
XX ABK42245;  
XX  
DT 21-MAY-2002 (first entry)  
XX  
DE Genomic sequence #144 encoding novel human connective tissue polypeptide.  
XX  
KW Human; connective tissue related disorder; cancer; gene therapy;  
XX  
OS Homo sapiens.  
XX  
PN WO20015343-A1.  
XX  
PD 02-AUG-2001.  
XX  
PE 17-JAN-2001; 2001WO-US01322.  
XX  
PF 31-JAN-2000; 2000US-0179065.  
XX  
PR 04-FEB-2000; 2000US-0180628.  
XX  
PR 24-FEB-2000; 2000US-0184664.  
XX  
PR 02-MAR-2000; 2000US-0186350.  
XX  
PR 16-MAR-2000; 2000US-0189874.  
XX  
PR 17-MAR-2000; 2000US-0190076.  
XX  
PR 18-APR-2000; 2000US-0198123.  
XX  
PR 19-MAY-2000; 2000US-0205515.  
XX  
PR 07-JUN-2000; 2000US-0209467.  
XX  
PR 28-JUN-2000; 2000US-0214886.  
XX  
PR 30-JUN-2000; 2000US-0215135.  
XX  
PR 07-JUL-2000; 2000US-0216647.  
XX  
PR 11-JUL-2000; 2000US-0216880.  
XX  
PR 11-JUL-2000; 2000US-0217487.  
XX  
PR 11-JUL-2000; 2000US-0218290.  
XX  
PR 14-JUL-2000; 2000US-0218290.  
XX  
PR 26-JUL-2000; 2000US-0220963.  
XX  
PR 26-JUL-2000; 2000US-0220964.  
XX  
PR 14-AUG-2000; 2000US-0224518.  
XX  
PR 14-AUG-2000; 2000US-0224519.  
XX  
PR 14-AUG-2000; 2000US-0225213.  
XX  
PR 14-AUG-2000; 2000US-0225214.  
XX  
PR 14-AUG-2000; 2000US-0225266.  
XX  
PR 14-AUG-2000; 2000US-0225267.  
XX  
PR 14-AUG-2000; 2000US-0225268.  
XX  
PR 14-AUG-2000; 2000US-0225270.  
XX  
PR 14-AUG-2000; 2000US-0225447.  
XX  
PR 14-AUG-2000; 2000US-0225757.  
XX  
PR 14-AUG-2000; 2000US-0225758.  
XX  
PR 14-AUG-2000; 2000US-0225759.  
XX  
PR 18-AUG-2000; 2000US-0226279.  
XX  
PR 22-AUG-2000; 2000US-0226681.  
XX  
PR 22-AUG-2000; 2000US-0226681.  
XX  
PR 22-AUG-2000; 2000US-0227182.  
XX  
PR 23-AUG-2000; 2000US-0227009.  
XX  
PR 30-AUG-2000; 2000US-0228924.







GenCore version 5.1.4-p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 23, 2003, 15:14:42 ; Search time 8.18182 Seconds  
(Without alignments)  
2938.192 Million cell updates/sec

Title: US-09-873-546-3  
Perfect score: 31  
Sequence: 1 gcggaattccacatgaggtgcatgtgcc 31

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 538826 seqs, 387737923 residues

Total number of hits satisfying chosen parameters: 1077652

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Published\_Applications\_NA: \*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEM\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEM\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEM\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEM\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_NEM\_PUB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US10\_NEM\_PUB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US60\_NEM\_PUB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19.2	61.9	340	10	US-09-880-107-2006
2	19	61.3	12932	10	US-09-764-847-1132
3	18.4	59.4	2936	10	US-09-747-835A-112
4	18.4	59.4	6663	10	US-09-037-657-28
5	18.4	59.4	11832	10	US-09-037-657-38
6	18.2	58.7	470	10	US-09-969-708-561
7	18.2	58.7	470	10	US-09-880-107-3150
8	18.2	58.7	661	10	US-09-770-149-464
9	18	58.1	7653	9	US-09-950-634-1
10	17.8	57.4	292	10	US-09-783-590-2469
11	17.8	57.4	296	9	US-10-040-739-549
12	17.8	57.4	514	10	US-09-864-761-12738
13	17.8	57.4	3666	9	US-09-895-913A-245
14	17.8	57.4	31718	9	US-09-764-872-812
15	17.8	57.4	31718	9	US-09-764-872-813
16	17.4	56.1	269	10	US-09-878-574-9123
17	17.4	56.1	485	10	US-09-815-242-2683
18	17.4	56.1	705	10	US-09-815-242-8867
19	17.4	56.1	705	10	US-09-815-242-8941

20	17.4	56.1	1539	9	US-09-938-842A-176	Sequence 176, App
21	17.4	56.1	32193	10	US-09-764-877-2623	Sequence 2623, Ap
22	17.2	55.5	466	10	US-09-864-761-5439	Sequence 5439, Ap
23	17.2	55.5	716	10	US-09-070-927A-866	Sequence 866, App
24	17.2	55.5	1345	10	US-09-764-864-315	Sequence 315, App
25	17.2	55.5	1841	10	US-09-764-864-735	Sequence 735, App
26	17.2	55.5	3633	12	US-10-044-090-123	Sequence 123, App
27	17.2	55.5	6633	10	US-09-995-542-1	Sequence 1, Appli
28	17.2	55.5	7621	10	US-09-954-456-307	Sequence 307, App
29	17.2	55.5	7621	10	US-09-954-456-2209	Sequence 2209, Ap
30	17	54.8	390	10	US-09-949-889-2	Sequence 2, Appli
31	17	54.8	1017	9	US-09-938-842A-60	Sequence 60, Appl
32	17	54.8	1710	9	US-09-912-628-2	Sequence 2, Appli
33	17	54.8	1743	9	US-10-006-856A-235	Sequence 235, App
34	17	54.8	1743	9	US-10-184-644-451	Sequence 451, App
35	17	54.8	7838	10	US-09-761-466-4	Sequence 4, Appli
36	17	54.8	148567	9	US-10-254-869-3	Sequence 3, Appli
37	17	54.8	148567	10	US-09-801-876B-3	Sequence 3, Appli
38	16.8	54.2	286	9	US-10-040-739-75	Sequence 75, Appl
39	16.8	54.2	362	10	US-09-878-574-1089	Sequence 1089, Ap
40	16.8	54.2	365	9	US-09-756-854-6	Sequence 6, Appli
41	16.8	54.2	365	9	US-10-041-574-6	Sequence 6, Appli
42	16.8	54.2	400	10	US-09-864-761-10344	Sequence 10344, A
43	16.8	54.2	410	10	US-09-967-768A-88	Sequence 88, Appl
44	16.8	54.2	420	10	US-09-864-761-23266	Sequence 23266, A
45	16.8	54.2	444	10	US-09-960-352-5693	Sequence 5693, Ap

#### ALIGNMENTS

RESULT 1  
US-09-880-107-2006  
Sequence 2006, Application US/09880107  
Patent No. US20020142981A1  
GENERAL INFORMATION:  
APPLICANT: Horne, Darci T.  
APPLICANT: Vockley, Joseph G.  
APPLICANT: Scherf, Uwe  
APPLICANT: Gene Logic, Inc.  
TITLE OR INVENTION: Gene Expression Profiles in Liver Cancer  
FILE REFERENCE: 44921-5028-MO  
CURRENT APPLICATION NUMBER: US/09/880,107  
CURRENT FILING DATE: 2001-06-14  
PRIOR APPLICATION NUMBER: US 60/211,379  
PRIOR FILING DATE: 2000-06-14  
PRIOR APPLICATION NUMBER: US 60/237,054  
PRIOR FILING DATE: 2000-10-02  
NUMBER OF SEQ ID NOS: 3950  
SOFTWARE: Patentln Ver. 2.1  
SEQ ID NO 2006  
LENGTH: 340  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: Genbank Accession No. US20020142981A1 H81070  
NAME/KEY: unsure  
LOCATION: (1) (340)  
OTHER INFORMATION: n = a or c or g or t  
US-09-880-107-2006  
Query Match 61.9%; Score 19.2; DB 10; Length 340;  
Best Local Similarity 87.5%; Pred. No. 13;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 7 TTCACATGAGGCGCATTTGCC 30  
Db 238 TTCCTGACGAGGCGCATTTGCC 261  
RESULT 2  
US-09-764-847-1132/c  
Sequence 1132, Application US/09764847



LENGTH: 11832  
TYPE: DNA  
ORGANISM: Unknown  
FEATURE:  
OTHER INFORMATION: Description of Unknown Organism: Murine NR6 gene  
US-09-037-657-38

Query Match  
Best Local Similarity 59.4%; Score 18.4; DB 10; Length 11832;  
78.6%; Pred. No. 56;  
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 3 GGATTTCACATGAGGTCATTGGCC 30  
||||| ||||| ||||| ||||| |||||  
DB 5765 GGATTTCACATGAGGTCATTGGCC 5738

RESULT 6  
US-09-969-708-561

Sequence 561, Application US/09969708  
Patent No. US20020102532A1

GENERAL INFORMATION:

APPLICANT: Augustus, Meena

TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu

FILE REFERENCE: 689290-70

CURRENT APPLICATION NUMBER: US/09/969,708

CURRENT FILING DATE: 2001-10-03

PRIOR APPLICATION NUMBER: US/60/237,606

PRIOR FILING DATE: 2000-10-03

PRIOR APPLICATION NUMBER: US/60/237,608

PRIOR FILING DATE: 2000-10-03

PRIOR APPLICATION NUMBER: US/60/237,425

PRIOR FILING DATE: 2000-10-03

NUMBER OF SEQ ID NOS: 658

SOFTWARE: PatentIn version 3.0

SEQ ID NO 561

LENGTH: 470

TYPE: DNA

ORGANISM: Homosapiens

FEATURE:

NAME/KEY: misc\_feature

OTHER INFORMATION: n-a, l, g or c

US-09-969-708-561

Query Match

Best Local Similarity 58.7%; Score 18.2; DB 10; Length 470;

80.0%; Pred. No. 40;

Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

7 TTCTCAGATGAGGTCATTGGCC 31

||||| ||||| ||||| ||||| |||||

DB 233 TTNTTGCAGAGGTCATTGGCC 257

RESULT 7

US-09-880-107-3150

Sequence 3150, Application US/09880107

Patent No. US20020142981A1

GENERAL INFORMATION:

APPLICANT: Horne, Darci T.

APPLICANT: Vockley, Joseph G.

APPLICANT: Scherf, Uwe

APPLICANT: Gene Logic, Inc.

TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer

FILE REFERENCE: 44921-5028-WO

CURRENT APPLICATION NUMBER: US/09/880,107

CURRENT FILING DATE: 2001-06-14

PRIOR APPLICATION NUMBER: US 60/211,379

PRIOR FILING DATE: 2000-06-14

PRIOR APPLICATION NUMBER: US 60/237,054

PRIOR FILING DATE: 2000-10-02

NUMBER OF SEQ ID NOS: 3950

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 3150

LENGTH: 470  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: Genbank Accession No. US20020142981A1 T56281  
US-09-880-107-3150

Query Match  
Best Local Similarity 58.7%; Score 18.2; DB 10; Length 470;  
80.0%; Pred. No. 40;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 7 TTCTCAGATGAGGTCATTGGCC 31  
||||| ||||| ||||| ||||| |||||

DB 233 TTNTTGCAGAGGTCATTGGCC 257

RESULT 8  
US-09-770-149-464/C

Sequence 464, Application US/09770149  
Patent No. US20020059663A1

GENERAL INFORMATION:

APPLICANT: Gorlach, Jorn

APPLICANT: An, Yong-Qiang

APPLICANT: Hamilton, Carol M.

APPLICANT: Price, Jennifer L.

APPLICANT: Raines, Tracy M.

APPLICANT: Yu, Yang

APPLICANT: Rameaka, Joshua G.

APPLICANT: Page, Amy

APPLICANT: Matthew, Abraham V.

APPLICANT: Ledford, Brooke L.

APPLICANT: Woessner, Jeffrey P.

APPLICANT: Haas, William David

APPLICANT: Garcia, Carlos A.

APPLICANT: Krickler, Maja

APPLICANT: Slader, Ted

APPLICANT: Davis, Keith R.

APPLICANT: Allen, Keith

APPLICANT: Hoffman, Neil

APPLICANT: Hurban, Patrick

TITLE OF INVENTION: Expressed Sequences of Arabidopsis

FILE REFERENCE: 2024 (PARA-013PRV)

CURRENT APPLICATION NUMBER: US/09/770,149

CURRENT FILING DATE: 2001-01-26

PRIOR APPLICATION NUMBER: 60/178,506

PRIOR FILING DATE: 2000-01-27

NUMBER OF SEQ ID NOS: 999

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 464

LENGTH: 661

TYPE: DNA

ORGANISM: Arabidopsis thaliana

US-09-770-149-464

Query Match

Best Local Similarity 58.7%; Score 18.2; DB 10; Length 661;

87.0%; Pred. No. 43;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

9 CTCACATGAGGTCATTGGCC 31

||||| ||||| ||||| ||||| |||||

DB 391 CTCACATGAGGTCATTGGCC 369

RESULT 9

US-09-950-634-1/C

Sequence 1, Application US/09950634

Publication No. US20030032775A1

GENERAL INFORMATION:

APPLICANT: Molnar-Kimber, Katherine L.

```

1      Falli, Amedeo F.
2      Cagliano, Thomas J.
3      Nakanishi, Koji
4      Chen, Yangtzu
5      TITLE OF INVENTION: EFFECTOR PROTEINS OF RAPAMYCIN
6      NUMBER OF SEQUENCES: 23
7      CORRESPONDENCE ADDRESS:
8      ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
9      Dunnet, L.L.P.
10     STREET: 1300 I Street, N.W.
11     CITY: Washington
12     STATE: DC
13     COUNTRY: USA
14     ZIP: 20005-3315
15     COMPUTER READABLE FORM:
16     MEDIUM TYPE: floppy disk
17     COMPUTER: IBM PC compatible
18     OPERATING SYSTEM: PC-DOS/MS-DOS
19     SOFTWARE: PatentIn Release #1.0, Version #1.30
20     CURRENT APPLICATION DATA:
21     APPLICATION NUMBER: US/09/950,654
22     FILING DATE: 13-Sep-2001
23     CLASSIFICATION: <unknown>
24     PRIOR APPLICATION DATA:
25     APPLICATION NUMBER: US 08/471,112
26     FILING DATE: 06-JUN-1995
27     APPLICATION NUMBER: US 08/384,524
28     FILING DATE: 13-FEB-1995
29     APPLICATION NUMBER: US 08/312,023
30     FILING DATE: 26-SEP-1995
31     APPLICATION NUMBER: US 08/207,975
32     FILING DATE: 08-MAR-1994
33     ATTORNEY/AGENT INFORMATION:
34     NAME: Slekman, Michael T.
35     REGISTRATION NUMBER: 56,276
36     REFERENCE/DOCKET NUMBER: 01142, 0058-00000
37     TELECOMMUNICATION INFORMATION:
38     TELEPHONE: 202-408-4000
39     TELEFAX: 202-408-4400
40     INFORMATION FOR SEQ ID NO: 1:
41     SEQUENCE CHARACTERISTICS:
42     LENGTH: 7653 base pairs
43     TYPE: nucleic acid
44     STRANDEDNESS: double
45     TOPOLOGY: linear
46     MOLECULE TYPE: cDNA TO MRNA
47     SEQUENCE DESCRIPTION: SEQ ID NO: 1:
48     US-09-950-634-1
49     Query Match          58.1%; Score 18; DB 9; Length 7653;
50     Best Local Similarity 80.8%; Fred. No. 80;
51     Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0.
52     QY      2 CGGAATTCACATGAGGTGCATT 27
53             | ||||| ||||| ||||| |||||
54     Db      3128 CAGAATTCCTCATGAGGTGACTAT 3103
55
56 RESULT 10
57 US-09-783-590-2469/c
58 : Sequence 2469, Application US/09783590
59 : Patent No. US20020110850A1
60 : GENERAL INFORMATION:
61 : APPLICANT: Dillon, Patrick J.
62 : APPLICANT: Haseltine, William A.
63 : APPLICANT: Li, Haodong
64 : APPLICANT: Rosen, Craig A.
65 : APPLICANT: Ruden, Steven M.
66 : TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
67 : FILE REFERENCE: PO-16,2C1
68 : CURRENT APPLICATION NUMBER: US/09/783,590
69 : CURRENT FILING DATE: 2000-02-15
70 : PRIOR APPLICATION NUMBER: 08/420,856

```

```

PRIOR FILING DATE: 1995-04-12
PRIOR APPLICATION NUMBER: 08/346,731
PRIOR FILING DATE: 1994-11-21
NUMBER OF SEQ ID NOS: 12465
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2469
LENGTH: 292
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (10)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (59)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (153)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (179)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (220)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (226)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (246)
OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-2469

Query Match          57.4%; Score 17.8; DB 10; Length 292;
Best Local Similarity 90.5%; Pred. No. 57;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps

OY      11 CACATGAGGTGCATTGGCC 31
        ||||| | |||||||
DB       287 CACAGCATGTGCATTGCC 267

RESULT 11
US-10-040-739-549
Sequence 549, Application US/10040739
Patent No. US20020173635A1
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
McCoy, John
Lavallee, Edward
Racie, Lisa
Merberg, David
Trecay, Maurice
Spaulding, Vikki
TITLE OF INVENTION: SECRETED, EXPRESSED SEQUENCE TAGS
NUMBER OF SEQUENCES: 1519
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
APPLICATION NUMBER: US/10/040,739
FILING DATE: 07-Jan-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
```

APPLICATION NUMBER: 09/036,520  
FILING DATE: 03-JUN-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: BROWN, Scott A.  
REGISTRATION NUMBER: 32,724  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 498-8224  
TELEFAX: (617) 876-5851  
INFORMATION FOR SEQ ID NO: 549:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 296 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
SEQUENCE DESCRIPTION: SEQ ID NO: 549:  
US-10-040-739-549

Query Match 57.4%; Score 17.8; DB 9; Length 296;  
Best Local Similarity 75.9%; Pred. No. 57;  
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
QY 1 GCGGAATTCACATGAGGCGCATTTGCC 29  
DB 68 GAGGACTTCCACAGAGGCGCATCTGCC 96

RESULT 12  
US-09-864-761-12738  
Sequence 12738, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aemica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
TITLE OF INVENTION: Identification of Polynucleotides  
FILE REFERENCE: 06132/043002  
CURRENT APPLICATION NUMBER: US/09/895,913A  
CURRENT FILING DATE: 2001-06-29  
PRIOR FILING DATE: 2000-02-04  
PRIOR FILING DATE: 2000-05-26  
PRIOR FILING DATE: 2000-08-03  
PRIOR FILING DATE: 2000-08-03  
PRIOR FILING DATE: 2000-10-04  
PRIOR FILING DATE: 2000-09-27  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21

PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 12738  
LENGTH: 514  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AL161747.1  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.91  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.83  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.82  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.99  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.54  
US-09-864-761-12738

Query Match 57.4%; Score 17.8; DB 10; Length 514;  
Best Local Similarity 75.9%; Pred. No. 63;  
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
QY 3 GGAATTCACATGAGGCGCATTTGCC 31  
DB 114 GGAGTACTCATTGAGCTTGCCTTGC 142

RESULT 13  
US-09-895-913A-245/c  
Sequence 245, Application US/09895913A  
Patent No. US20020160456A1  
GENERAL INFORMATION:  
APPLICANT: Kleantous, Harold  
APPLICANT: Al-Garawi, Amal  
APPLICANT: Miller, Charles  
APPLICANT: Tomb, Jean Francois  
APPLICANT: Oomen, Raymond P.  
TITLE OF INVENTION: Identification of Polynucleotides  
FILE REFERENCE: Encoding No. US20020160456A1el Helicobacter Polypeptides in t  
FILE REFERENCE: 06132/043002  
CURRENT APPLICATION NUMBER: US/09/895,913A  
CURRENT FILING DATE: 2001-06-29  
PRIOR FILING DATE: 2001-06-29  
PRIOR FILING DATE: 2001-06-24  
NUMBER OF SEQ ID NOS: 368  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 245  
LENGTH: 3666  
TYPE: DNA  
ORGANISM: Helicobacter pylori  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (54)...(3608)  
US-09-895-913A-245

Query Match 57.4%; Score 17.8; DB 9; Length 3666;  
Best Local Similarity 75.9%; Pred. No. 88;  
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
QY 3 GGAATTCACATGAGGCGCATTTGCC 31  
DB 364 GGCAGCGCTCATATCAGGAGCATTTGCC 336

RESULT 14  
US-09-764-872-812  
Sequence 812, Application US/09764872  
Patent No. US20030050231A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

```

; FILE REFERENCE: PA125
; CURRENT APPLICATION NUMBER: US/09/764,872
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 957
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 812
; LENGTH: 31718
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-872-812

```

```

Query Match          57.4%; Score 17.8; DB 9; Length 31718;
Best Local Similarity 75.9%; Pred. No. 1.3e+02;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

```

OY      1 GCGGAATTCTCACATGAGGGTGCAATTGC 29
DB 24277 GAGGACTTCCACAGAGAGGTGGCATCTGC 24305

```

```

RESULT 15
US-09-764-872-813
; Sequence 813, Application US/09/764872
; Publication No. US20030050231A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA125
; CURRENT APPLICATION NUMBER: US/09/764,872
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 957
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 813
; LENGTH: 31718
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-872-813

```

```

Query Match          57.4%; Score 17.8; DB 9; Length 31718;
Best Local Similarity 75.9%; Pred. No. 1.3e+02;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

```

OY      1 GCGGAATTCTCACATGAGGGTGCAATTGC 29
DB 24277 GAGGACTTCCACAGAGAGGTGGCATCTGC 24305

```

```

Search completed: March 23, 2003, 17:17:01
Job time : 29.1818 secs

```





BASE COUNT	109 a	103 c	111 g	97 t
ORIGIN	/note="Callus K19"			
Query Match	74.2%	Score 23;	DB 9;	Length 420;
Best Local Similarity	83.9%;	Pred. No. 16;		
Matches 26;	Conservative 0;	Mismatches 5;	Indels 0;	Gaps 0;
OY	1	GGGAAATTCACATGAGGCGTCATTGGCC	31	
DB	255	GCAGAAATTTCCCATGAGCGTCGTTTGCC	285	
RESULT 2				
AV835621/c				
LOCUS	AV835621	574 bp	MRNA	linear
DEFINITION	AV835621 K. Sato unpublished cDNA library: Hordeum vulgare subsp. spontaneum top three leaves adult, heading stage Hordeum vulgare subsp. spontaneum cDNA clone bah3110, mRNA sequence.			
ACCESSION	AV835621			
VERSION	AV835621.1	GI:14527710		
KEYWORDS	EST.			
SOURCE	Hordeum vulgare subsp. spontaneum.			
ORGANISM	Hordeum vulgare subsp. spontaneum			
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae; Triticeae; Hordeum.			
AUTHORS	1 (bases 1 to 574)			
TITLE	Sato, K.			
JOURNAL	Barley EST sequencing project in NIG and Okayama Univ Unpublished (2001)			
COMMENT	Contact: Kazuhiro Sato Research Institute for Bioreources Okayama University, Barley Germplasm Center Chuo 2-20-1, Kurashiki, Okayama 710-0046, Japan Email: kazatco@rib.okayama-u.ac.jp, URL: http://www.rib.okayama-u.ac.jp/barley/ database: http://www.shigen.nig.ac.jp/barley/html. Location/Qualifiers			
FEATURES	1..574			
source	/organism="Hordeum vulgare subsp. spontaneum" /cultivar="H602" /db_xref="taxon:77009" /clone="bah3110" /clone_1lb="K. Sato unpublished cDNA library: Hordeum vulgare subsp. spontaneum top three leaves adult, heading stage" /issue_type="top three leaves" /tissue_type="adult, heading stage" /dev_stage="adult, heading stage"			
BASE COUNT	138 a	150 c	138 g	146 t
ORIGIN	2 others			
Query Match	74.2%	Score 23;	DB 10;	Length 574;
Best Local Similarity	83.9%;	Pred. No. 17;		
Matches 26;	Conservative 0;	Mismatches 5;	Indels 0;	Gaps 0;
OY	1	GGGAAATTCACATGAGGCGTCATTGGCC	31	
DB	216	GCAGAAATTTCCCATGAGCGTCGTTTGCC	186	
RESULT 3				
BE885242/c				
LOCUS	BE885242	1119 bp	MRNA	linear
DEFINITION	BE885242 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3907851 5', mRNA sequence.			
ACCESSION	BE885242			
VERSION	BE885242.1	GI:10334018		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			

REFERENCE	1 (bases 1 to 1119)
AUTHORS	NIH-MGC <a href="http://mgc.nhl.nih.gov/">http://mgc.nhl.nih.gov/</a> .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:cgapbs-r@mail.nih.gov">cgapbs-r@mail.nih.gov</a> Tissue Procurement: ATCC cDNA Library Preparation: Life Technologies, Inc. DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Plate: LHM9719 row a column: 04 High quality sequence stop: 231. Location/Qualifiers
FEATURES	1..1119
source	/organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:3907851" /clone_11b="NIH-MGC-71" /tissue_type="TelomYosarcoma" /lab_host="DH10B (phage-resistant)" /note="Organ: uterus; Vector: pCMV-SPORT6; Site: 1; Notif: Site 2: Salt: Cloned unidirectionally. Primer: Oligo dt. Average insert size 2.1 kb."
BASE COUNT	481 a 291 c 305 g 42 t
ORIGIN	
Query Match	74.2%; Score 23; DB 12; Length 1119;
Best Local Similarity	100.0%; Pred. No. 19;
Matches	23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	9 CTCACATGAGGGGTGCATTGGCCC 31 
Db	206 CTCACATGAGGGGTGCATTGGCCC 184
LOCUS	A0349812
DEFINITION	757 bp DNA linear GSS 07-MAY-1999
ACCESSION	A0349812
VERSION	A0349812.1 GI:4177147
KEYWORDS	GSS.
SOURCE	human.
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 757) Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter,J.C.
TITLE	Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building
JOURNAL	Unpublished (1997)
COMMENT	Other-GSSs: RPCI11-118N20.TV Contact: Shaying Zhao, William Nierman, Mark Adams Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850 Tel: 301 838 0200 Fax: 301 838 0208 Email: <a href="mailto:hbe@ligr.org">hbe@ligr.org</a> Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong <a href="mailto:pieter@ligr.org">pieter@ligr.org</a> ( <a href="mailto:med.bu@ligr.org">med.bu@ligr.org</a> ). Clones may be purchased from BACPAC Resources ( <a href="http://bacpac.med.bu@ligr.org">http://bacpac.med.bu@ligr.org</a> ) or from Research Genetics ( <a href="http://info@resgen.com">info@resgen.com</a> ). BAC end search page: <a href="http://www.ligr.org/tldb/humgen/bac_end_search/bac_end_search.html">http://www.ligr.org/tldb/humgen/bac_end_search/bac_end_search.html</a> Seq primer: SP6 Class: BAC ends. Location/Qualifiers
FEATURES	

```

source
1. 757
/organism="Homo sapiens"
/db_xref="GDB:7545259"
/db_xref="taxon:9606"
/clone="RPC1-11-118N20"
/clone_id="RPC1-11"
/sex="Male"
/cell_type="Lymphocytes"
/notes="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
RPC11 Human Male BAC library"
BASE COUNT      206 a      166 c      137 g      246 t      2 others
ORIGIN
Query Match      68.4%; Score 21.2; DB 17; Length 757;
Best Local Similarity 88.5%; Pred. No. 1.1e+02;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 GCGGAATTCACATGAGGTCGATC 26
|||||
Db 6 GCGGAATTCACATGAGGTCGATC 31

RESULT 5
BF796120/c      984 bp      mRNA      linear      EST 12-JAN-2001
LOCUS           602258969F1 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:4342360 5',
DEFINITION      mRNA sequence.
ACCESSION       BF796120
VERSION         BF796120.1 GI:12101174
KEYWORDS        EST.
SOURCE          human.
ORGANISM        Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE       1 (bases 1 to 984)
AUTHORS         NIH-MGC http://mgc.nci.nih.gov/.
TITLE           National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL         Unpublished (1999)
COMMENT         Contact: Robert Strausberg, Ph.D.
                 Email: cgabs@email.nih.gov
                 Tissue Procurement: Louis Staudt, M.D., Ph.D.
                 cDNA Library Preparation: Life Technologies, Inc.
                 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
                 DNA Sequencing by: Incyte Genomics, Inc.
                 Clone distribution: MGC clone distribution information can be
                 found through the I.M.A.G.E. Consortium/LNLN at:
                 http://image.llnl.gov
                 Plate: LHAM957 row: 1 column: 17
                 High quality sequence start: 13
                 High quality sequence stop: 628.
FEATURES
source
1..984
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="IMAGE:4342360"
/clone_lib="NIH_MGC_85"
/tissue_type="Lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: lymph; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.867 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC library."
BASE COUNT      248 a      263 c      290 g      183 t
ORIGIN
Query Match      67.7%; Score 21; DB 12; Length 984;
Best Local Similarity 82.8%; Pred. No. 1.4e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 3 GGAATTCACATGAGGTCGATTCGCC 31
|||||
Db 870 GGTATTCCTCCATGGGGGTGATTCGCC 842

```

```

RESULT 6
BG955224
LOCUS           385 bp      mRNA      linear      EST 12-JUN-2001
DEFINITION      CM4-CT0657-120201-861-910 CT0657 Homo sapiens cDNA, mRNA sequence.
ACCESSION       BG955224
VERSION         BG955224.1 GI:14373395
KEYWORDS        EST.
SOURCE          human.
ORGANISM        Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE       1 (bases 1 to 385)
AUTHORS         Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE           Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL         Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE        20202663
COMMENT         Contact: Simpson A.J.G.
                 Laboratory of Cancer Genetics
                 Ludwig Institute for Cancer Research
                 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
                 Brazil
                 Tel.: +55-11-2704922
                 Fax: +55-11-2707001
                 Email: asimpson@ludwig.org.br
                 This sequence was derived from the FAPESP/LICR Human Cancer Genome
                 Project. This entry can be seen in the following URL
                 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM4ct2-CM4-CT0657-
                 120201-861-g106t3-2001-02-126t4-1)
                 Seq primer: puc 18 forward
                 High quality sequence start: 69
                 High quality sequence stop: 385.
FEATURES
source
1..385
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="CT0657"
/dev_stage="Adult"
/notes="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the puc 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT      97 a      55 c      68 g      165 t
ORIGIN
Query Match      66.5%; Score 20.6; DB 13; Length 385;
Best Local Similarity 85.2%; Pred. No. 1.7e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 4 GAATTCACATGAGGTCGATTCGCC 30
|||||
Db 220 GAATTCACATGAGGTCGATTCGCC 246

RESULT 7
AM128188/c
LOCUS           528 bp      mRNA      linear      EST 25-OCT-1999
DEFINITION      f111f07.x1 Sugano Kawakami zebrafish DRF Danilo ref10 cDNA clone
2600869 3' similar to contains element MER30 repetitive element ;,
mRNA sequence.
ACCESSION       AM128188
VERSION         AM128188.1 GI:6116122
KEYWORDS        EST.

```

SOURCE zebrafish.  
ORGANISM Danio rerio  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.  
REFERENCE 1 (bases 1 to 528)  
AUTHORS Sugano, S., Kawakami, K., Johnson, S., Li, F., Marra, M., Eddy, S., Hillier, L., Clifton, S., Allen, M., Gibbs, M., Jost, S., Kucaba, T., Martin, J., Pape, D., Stepien, M., Underwood, K., Theising, B., Riltter, E., Bowers, Y., Wylie, T., Waterston, R. and Wilson, R.  
TITLE WashU Zebrafish EST Project 1999  
JOURNAL Unpublished (1999)  
COMMENT Contact: S.L. Johnson  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA Sequencing by: Washington University Genome Sequencing Center  
Seq primer: T7 ET from Amersham  
High quality sequence stop: 479.  
Location/Qualifiers  
1..528  
/organism="Danio rerio"  
/strain="AB"  
/db\_xref="taxon:7955"  
/clone="2600869"  
/clone\_1lb="Sugano Kawakami zebrafish DNA"  
/sex="mixed (one male and one female, including unfertilized eggs)"  
/dev\_stage="adult"  
/lab\_host="DH10B (phage resistant)"  
/note="vector: pME18S-FL3; Site\_1: DraIII (CACTGTG); Site\_2: DraIII (CACCATGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGCGCCCTTTTCTTTTCTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTGGCCCTACTG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science) and kindly donated by Dr. Koichi Kawakami. Custom primers for sequencing: 5' end primer CTTCTCTCTTAAGCTGCG and 3' end primer CGACCTGCAGCTCGACACA."

BASE COUNT 193 a 92 c 74 g 169 t  
ORIGIN

Query Match 66.5%; Score 20.6; DB 10; Length 528;  
Best Local Similarity 85.2%; Pred. No. 1.8e+02;  
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 GGAATTCACATGAGGTCGATTCG 29  
||||| ||||| ||||| ||||| |||||  
DB 503 GGAATTCACATGAGGTCGATTCG 477

RESULT 8  
LOCUS A1226051  
DEFINITION u108f10.y1 Sugano mouse liver m1a mus musculus cDNA clone  
IMAGE:1891339.5' similar to gb:J04973 UBIOUINOL-CTTCHROME C  
REDUCTASE CORE PROTEIN 2 PRECURSOR (HUMAN); mRNA sequence.  
ACCESSION A1226051.1 GI:3809104  
VERSION A1226051.1  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 704)  
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Scheinberg, K., Stepien, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.  
TITLE The WashU-HMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT Contact: Maria M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:975663  
Trace considered overall poor quality  
Seq primer: custom primer used  
High quality sequence stop: 1.  
Location/Qualifiers  
1..704  
/organism="Mus musculus"  
/strain="C57BL"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1891339"  
/clone\_1lb="Sugano mouse liver m1a"  
/sex="female"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="organ: liver; Vector: pME18S-FL3; Site\_1: DraIII (CACTGTG); Site\_2: DraIII (CACCATGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGCGCCCTTTTCTTTTCTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTGGCCCTACTG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTTCTCTCTTAAGCTGCG and 3' end primer CGACCTGCAGCTCGACACA."

BASE COUNT 190 a 158 c 172 g 184 t  
ORIGIN

Query Match 65.8%; Score 20.4; DB 9; Length 704;  
Best Local Similarity 95.5%; Pred. No. 2.3e+02;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 AATTCACATGAGGTCGATT 26  
||||| ||||| ||||| ||||| |||||  
DB 156 AATTCACATGAGGTCGATT 177

RESULT 9  
LOCUS B1121740  
DEFINITION F045P75Y Populus flower cDNA library Populus balsamifera subsp.  
trichocarpa cDNA, mRNA sequence.  
ACCESSION B1121740  
VERSION B1121740  
KEYWORDS EST.  
SOURCE Populus balsamifera subsp. trichocarpa.  
ORGANISM Populus balsamifera subsp. trichocarpa  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eustoids I; Malpighiales; Salicaceae; Populus.  
REFERENCE 1 (bases 1 to 206)  
AUTHORS Hertzberg, M., Aspeberg, H., Erlandsson, R., Bjorkbacka, H., Hiltunen, T., Karlsson, J., Teerl, T., Gustafsson, P., Bahler, R., Jansson, S., Nilsson, O., Sundberg, B., Nilsson, P., Uhlen, M., Sandberg, G. and Lundberg, J.  
TITLE Gene expression in Populus



ACCESSION AQ378111  
 VERSION AQ378111.1 GI:4349134  
 KEYWORDS GSS.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 589)  
 AUTHORS Zhao, S., Adams, M.D., Nierman, W., Malek, J., de Jong, P. and Venter, J.C.  
 TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Shaying Zhao, William Nierman, Mark Adams  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: hbeetlgr.org  
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genetics (inforesgen.com). BAC end search page: http://www.tigr.org/tldb/hungen/bac\_end\_search/bac\_end\_search.html  
 Seq primer: SP6  
 Class: BAC ends.  
 FEATURES  
 source Location/Qualifiers  
 1..589  
 /organism="Homo sapiens"  
 /db\_xref="GDB:7561886"  
 /db\_xref="taxon:9606"  
 /clone="RPCI-11-162C15"  
 /clone\_lib="RPCI-11"  
 /sex="Male"  
 /cell\_type="Lymphocytes"  
 /note="Vector: pBAC3.6; Site\_1: EcoRI; Site\_2: EcoRI; RPCI11 Human Male BAC Library"  
 BASE COUNT 177 a 121 c 137 g 153 t 1 others  
 ORIGIN  
 Query Match 64.5%; Score 20; DB 17; Length 589;  
 Best Local Similarity 82.1%; Pred. No. 3.3e+02;  
 Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 Oy 1 GCGGAAATTCACATGAGGTCATTTG 28  
 ||||||| ||||||| ||| |||  
 b 169 GCGGAAATTCACATGAGGTCATTTG 196  
 ||||||| ||||||| ||| |||  
 RESULT 13  
 BI935904/c 661 bp mRNA linear EST 18-OCT-2001  
 LOCUS EST555793 tomato flower, anthesis Lycopersicon esculentum cDNA  
 DEFINITION clone cTOD24124 5' end, mRNA sequence.  
 ACCESSION BI935904  
 VERSION BI935904  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Lycopersicon esculentum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum;  
 Lycopersicon.  
 1 (bases 1 to 661)  
 van der Hoeven, R.S., Bezzerides, J.L., Karamycheva, S.A., Tsai, J., Uitterback, T., Van Aken, S., Rønning, C.M., Nierman, W., Fraser, C.M., Martin, G.B., Giovannoni, J.J. and Tanksley, S.D.  
 Generation of ESTs from tomato flower tissue, anthesis (2001)  
 TITLE Unpublished (2001)  
 JOURNAL  
 COMMENT Contact: CUGI  
 Clemson University Genomics Institute

Clemson University  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Email: http://www.genome.clemson.edu/Orders/index.html  
 This clone is available through the Clemson University Genomics Institute  
 Seq primer: T3.  
 FEATURES  
 source Location/Qualifiers  
 1..661  
 /organism="Lycopersicon esculentum"  
 /cultivar="TA496"  
 /db\_xref="taxon:4081"  
 /clone="cTOD24124"  
 /clone\_lib="tomato flower, anthesis"  
 /tissue\_type="flower"  
 /dev\_stage="anthesis"  
 /note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2: XhoI; supplier: Cornell University; sequencing: The Institute for Genomic Research; flower buds and flowers were taken from greenhouse plants (4-8 wks old, TA496). They were immediately frozen in liquid nitrogen and then size-separated while remaining frozen."  
 BASE COUNT 183 a 92 c 120 g 266 t  
 ORIGIN  
 Query Match 64.5%; Score 20; DB 13; Length 661;  
 Best Local Similarity 82.1%; Pred. No. 3.4e+02;  
 Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 Oy 4 GAATTCACATGAGGTCATTTGCC 31  
 ||| ||||||||| || |||  
 Db 281 GAATTCACATGAGGTCATTTCCC 254  
 ||| ||||||||| || |||  
 RESULT 14  
 CNS02FVQ/c 785 bp DNA linear GSS 13-MAY-2000  
 LOCUS CNS02FVQ/c  
 DEFINITION Tetradon nigroviridis genome survey sequence PUC-Orl end of clone 135M02 of library G from Tetradon nigroviridis, genomic survey sequence.  
 ACCESSION AL195551.1 GI:7833657  
 VERSION AL195551  
 KEYWORDS GSS: genome survey sequence.  
 SOURCE Tetradon nigroviridis.  
 ORGANISM Tetradon nigroviridis.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.  
 1 (bases 1 to 785)  
 Roest-Crolius, H., Jallou, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fitzames, C., Winkler, P., Brothier, P., Quetier, F., Saurin, W. and Weissenbach, J.  
 Human gene number estimate provided by genome wide analysis using Tetradon nigroviridis DNA sequence  
 UNPUBLISHED  
 2 (bases 1 to 785)  
 Roest-Crolius, H., Jallou, O., Dasilva, C., Fitzames, C., Fisher, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.  
 Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetradon nigroviridis  
 UNPUBLISHED  
 3 (bases 1 to 785)  
 Direct Submission  
 Submitted (12-APR-2000)  
 This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetradon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetradon.  
 TITLE  
 JOURNAL  
 COMMENT  
 FEATURES  
 source Location/Qualifiers  
 1..785  
 /organism="Tetradon nigroviridis"

/db\_xref="taxon:99883"  
/clone="135M02"  
/clone\_1lb="G"  
/note="Genoscope sequence ID : COAG135B01SP1-end :  
PUC-ori"

BASE COUNT 197 a 185 c 169 g 176 t 58 others

ORIGIN

Query Match 64.5%; Score 20; DB 17; Length 785;  
Best Local Similarity 82.1%; Pred. No. 3.5e+02;  
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 4 GAATTCACATGAGGTCATTTGCC 31  
|||||  
Db 207 GCATGCTCACATGAAGTTCATGTGCC 180

RESULT 15  
BQ719125 899 bp mRNA linear EST 16-JUL-2002  
LOCUS AGENCOURT\_8232632 lupski\_sympathetic\_trunk Homo sapiens cDNA clone  
DEFINITION IMAGE:6191580 5', mRNA sequence.  
ACCESSION BQ719125  
VERSION BQ719125.1 GI:21858022  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 899)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabs-remail.nih.gov  
Tissue Procurement: Dr. James R. Lupski  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.lnl.gov  
Plate: LLM13592 row: d column: 13  
High quality sequence stop: 678.  
Location/Qualifiers  
1. 899

FEATURES  
source

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6191580"  
/clone\_1lb="lupski\_sympathetic\_trunk"  
/sex="male"  
/tissue.type="sympathetic trunk"  
/dev\_stage="adult, 16 yr"  
/lab\_host="DH10B"  
/note="Vector: PCMV-SPORT6 (Life Technologies); Site\_1:  
NotI; Site\_2: SalI; cDNA made by oligo-dT priming.  
Directionally cloned using the following adaptors:  
5'-TCGACCCAGCGTCGC-3' and  
5'-GACTACTTCAGATCGGAGCGGCCCT(15)-3'. Size selected >  
1 kb for average insert length 1.9 kb. This is a primary  
library, non-amplified. Library constructed by Life  
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor  
College of Medicine); available through Life  
Technologies."

BASE COUNT 230 a 176 c 249 g 244 t

ORIGIN

Query Match 64.5%; Score 20; DB 14; Length 899;  
Best Local Similarity 82.1%; Pred. No. 3.6e+02;  
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 4 GAATTCACATGAGGTCATTTGCC 31  
|||||

Db 346 GAATTCACACAGGAGTGCATTTGCC 373  
Search completed: March 23, 2003, 17:12:25  
Job time : 206.545 secs

**THIS PAGE BLANK (USPTO)**